Vol. 43 OCTOBER, 1952 No. 6

The Journal of Educational Psychology

UNIVERSITY OF MICHIGAN
OCT 15 1952
PERIODICAL READING ROOM

CONTENTS

Learning—"They Ain't No Such Animal" 321
HORACE B. ENGLISH

Action Research by Teachers and the Population Sampling Problem 331
STEPHEN M. COREY

A Validation of Development and Adjustment Hypotheses of Adolescence 339
AILEEN SCHOEPPE AND R. J. HAVIGHURST

The Factor Analysis of Subject-matter Experts 354
WARREN S. TORGERSON AND BERT F. GREEN, JR.

Sequence in Multiple Choice Item Options 364
SCARVIA B. ANDERSON

Book Reviews 369

Publications Received 382

Published monthly except June to September

$7.00 a year in the United States, Pan American Countries and the Philippines; Canada, $7.20; other countries, $7.40. Single issues $1.20

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF Educational Psychology

Established 1910

BOARD OF EDITORS*

STEPHEN M. COREY, *Chairman*
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

WM. CLARK TROW
University of Michigan

* Appointed by the American Psychological Association.

EDITORIAL CONSULTANTS

JACK W. DUNLAP
KARL J. HOLZINGER
HAROLD E. JONES
H. H. REMMERS
PERCIVAL M. SYMONDS
MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. The JOURNAL will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

Manuscripts and communications regarding editorial matters should be addressed to the Chairman of the Board of Editors, Stephen M. Corey, Teachers College, Columbia University, New York 27, N. Y.

Books and other materials for review, and business communications should be addressed to THE JOURNAL OF EDUCATIONAL PSYCHOLOGY, Warwick & York, Publishers, 10 E. Centre St., Baltimore 2, Md.

Manuscripts should be typed and double-spaced throughout, including quotations, footnotes, and references. In order to attain

(Continued on Inside Back Cover)

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 43 October, 1952 Number 6

LEARNING—"THEY AIN'T NO SUCH ANIMAL."*

HORACE B. ENGLISH
The Ohio State University

It is only within the memory of men now living that psychologists began explicitly to postulate a specific and unitary construct by means of which to explain the universal fact that organisms learn. A theory of learning is a product of the very recent past.

The earlier psychologists, you remember, spoke of memory and of association. Baldwin's big two-volume *Dictionary of Psychology* in 1902 does not even list the term learning. Warren's *Dictionary* in 1934 gave learning a place but only as a term belonging to applied psychology. My own little *Dictionary* in the same year defined learning in terms of improvement—thus making it a technological concept—but also attempted to define it in scientific terms. It is not until the era of Thorndike—say around 1910—that learning came into widespread use as a basic psychological construct. Now, however, learning crowds association and even memory clear out of the indexes of our textbooks.

Yet the new construct did not spring full-grown from the Jovian brow of Thorndike like Athena from the head of Zeus. Learning came into psychology as the child of associationism and wittingly or unwittingly we all conceive learning in terms which we have inherited from that ancient doctrine.† Research on learning has been almost wholly dominated by the view that it

* A portion of the Presidential Address, the Division of Educational Psychology of the American Psychological Association, 1951.

† It is very interesting to trace the way in which associationism was shaped at each era by the kind of thinking prevalent in the scientific world. Hume sought laws of association to parallel the laws of gravitation; later association became mental chemistry; later still, telephony. Nowadays it is more acceptable or 'advanced' to conceive of association in 'field theoretical' terms. Except for the last view, which has not yet managed to alter the vocabulary and hence the thinking of the man in the street, all these views may be fairly characterized as 'atomistic.'

consists in acquiring and retaining something—usually a 'stimulus-response bond.' This research has, I believe, made notable contributions to the general theory of behavior, but, curiously enough, has given us almost no light on the specific problems of how organisms learn; that is, of how they come to improve their behavior.

There have been various proposals for correcting the narrowness of the associationist position. One of the earliest was set forth by McDougall and May Smith. It proposed a two-factor theory—one a fairly strict contiguity factor, the other concerned with the apprehension of meaning. I have carried out a considerable number of investigations designed to test this hypothesis. My experiments, though still inconclusive, include over a quarter-million distinct, measured responses.* The Gestalters and neo-Gestalters (e.g. Katona) have brought forward evidence for a kind of learning which cannot be reduced to contiguity-conditioning. And Tolman and more recently Mowrer have proposed bi- or multi-factor theories of learning.

But the search in all these cases was still for the process or processes which are specific to learning—i.e., for processes always present when learning is taking place, absent when learning is not taking place. Now it is my suggestion that there are no such processes and that therefore the construct of learning and of the learning process is scientifically invalid. Note that I do not deny that there are learning 'phenomena.' My point is that they belong together under one heading only on a purely formal basis.

All psychophysical or organic activity leads inevitably to a change in performance. You can call this change learning if you like, but to segregate the aspect of change from the totality of organic activity is an abstraction. When we perceive, when we think, when we emote, we certainly change in our capacity for later behavior; but the laws or explanatory principles are those of perceiving or thinking or emoting. There is no set of laws specific to the changing of performance, hence there is no valid construct of learning in the science of psychology.

---

* I cannot quite forbear remarking that I am unimpressed with experiments which measure only twenty or thirty responses in any given category, even when the found differences meet small-sample criteria of significance. Surely the day is not too far off when we shall recognize small-sample studies as pilot experiments only.

I would not on that account abandon all talk of learning, but I would restore to the term its original meaning. Try as we will, we cannot for long avoid bringing in the notion of learning as 'improvement.' Now that means that we leave the field of science for the field of psychotechnology, and that is where the concept of learning belongs.

In making this distinction I am not seeking to drive a broad wedge between psychology as science and as technology. The relationships and interactions between them are many and significant. For one thing, the whole broad range of scientific method is entirely common ground. Nonetheless the two disciplines have distinct goals and tend, therefore, to develop different concepts and working tools.

Science works with such a division or classification of all intrinsically related phenomena as will enable us to explain nature in the simplest possible terms (or, what comes to the same thing, in the fewest or most universal laws). This Principle of Parsimony is a basic regulator of all science. Technology, on the other hand, works with groupings of phenomena which have been segregated because their manipulation and control serves a human or social purpose. The phenomena brought together in a technology may have few intrinsic relations with each other; consider, for example, the variety of phenomena grouped together in oil geology!* Science, then, seeks parsimonious explanation, technology practical control.

It must be said at once that the distinction just made is not absolute: control over nature is the ultimate goal of even the purest science. And even the narrowest technology demands that its rules be as few and as inclusive as possible—within limits of the practical purposes from which it starts. But each practical purpose leads to its own set of rules, and these may overlap in complex fashion with the rules or working principles developed for other purposes. A tremendous number of rules and explanatory principles is thus generated for technological use. Few of these principles fit into the pattern of scientific law or even sug-

* We are here including applied science under the term technology. If desired, distinction may be made between the search for general principles with immediately practical ends in view (applied science), and the manipulation of a specific or concrete situation (technology). When pushed, however, the distinction appears one of degree.

gest these laws. Thus the rules for building a brick wall do not pointedly suggest the laws of physics and chemistry, although of course such laws apply. Nor could one derive the practical rules for building a wall from the law of gravitation and the chemistry of mortar.

A simple illustration will suffice to show how psychological technology may proceed in ways which cut across the lines of scientific theory. It is commonly accepted that persons suffering from 'feelings of inadequacy' tend to learn slowly. The technologist would wish to learn as exactly as possible the extent of this correlation—what kinds of things they learn slowly, for example. And he would seek to find methods of teaching which are adapted to different levels and areas of 'feelings of inadequacy.'* Now such an investigation is not very likely to add to psychological theory, though of course it might bring to light facts which in another context would magnificently serve theory. (The most unlikely-looking fact sometimes yields scientific illumination.) Certainly such investigation is very unlikely to add to our understanding of what 'learning really is.' Yet there are many concrete situations where classifying learners in terms of 'feelings of inadequacy' is of great help in the effort to produce the greatest degree of learning.

The frank recognition that 'learning' is a concept which deals with improvement, that is with practical ends, seems to promise considerable advantage both to the science of psychology, from which the term would be excluded, and to psychotechnology, to which the concept would be assigned. Take pure science first. Bringing together, under the one concept of learning, phenomena which do not belong together has led to over-generalization on the one hand and to a frustrating failure to find generalizations on the other. In the laboratory and in the study we have made valiant effort to find common principles of explanation for all the many and diverse personal activities which have been brought together under the single rubric of learning. So far we seem to have failed. Abandonment of that effort, at least until more is known

* We need not, for the purpose of this illustration, raise the question as to whether, even for technological purposes, the concept of 'inadequacy' is happily constructed. And of course we do not assume that feelings of inadequacy are in all cases unalterable barriers to learning, but merely that we may often have to proceed as if they were.

of the apparently distinct activities, will release the energies of investigators for the study of more specifically conceived problems, will lead to better experimental designs.

What are these problems? It would be presumptuous of me to attempt a complete list. Certainly, however, many of the problems studied as if they were learning problems are problems of perceptual discrimination; nothing is added by roundly proclaiming that their study is a contribution to learning theory. These problems grade over into problems of so-called insight, of conceptualization, of thinking and of judgment. Other problems are perhaps more appropriately categorized as those of motivation—at least this construct, while still confused, seems to have promise for the ordering of the data of our science. There are many other proposed conceptualizations of the facts of behavior —you need not fear that I shall attempt here and now to review them. What I am suggesting is that they have been less fruitfully attacked because conceived as subproblems of learning. As scientific problems they are more profitably attacked independently of any alleged relationship with a superordinate concept of learning. That has been in fact the actual procedure in many cases; it should be in all.

As these specific problems are vigorously attacked, we shall have a better chance to discover whether there is indeed an 'element' common to them all which justifies the construct of learning, and if so what its nature may be. The hypothesis that there is such a common process is currently beset by so many apparent discrepancies that it is, at the least, premature, and our attachment to this hypothesis is a hindrance to a clear envisagement of our actual problems in pure and systematic psychology. My suggestion is, therefore, that 'learning' is not a construct which brings order to the facts of psychology; rather it imposes a spurious, and in the end confusing, unity upon diverse phenomena.

Even if, however, it shall eventually prove that there is a proper place for the concept of learning in pure science, we must recognize the validity of considering learning as an independent psychotechnological problem. We have, as a matter of fact, always been insistent that learning had practical implications. The whole approach has been characteristically American, and as such influenced by the dominant characteristics of American scholarship, especially by a passion to be practical and an almost

fanatical devotion to facts. The student of learning, however much he designed his experiments as a venture in pure science using rats as his subjects, almost always had one eye cocked on practical conclusions. And he often found them.

But psychology is dominated also by another American characteristic, the yearning for a relatively simple and all-inclusive formula. And the effort to impose such an all-inclusive formula upon the data has tended to restrict our vision. Facts which do not fit the formula are too often 'filed and forgot.' Thus much of great importance which deals with the improvement of behavior is neglected.

When, however, we put aside the desire for the all-inclusive formula, and face learning as a technological concept, a whole nest of facts and problems comes to light. The variables and dimensions to be considered are seen as far more diversified than those of traditional learning experimentation. Thus, both the enduring and the temporary condition or state of the learner, instead of being ignored or treated as a source of constant or variable 'errors,' come to life as by far the most significant among the determiners of learning. The integrated organism ('The whole child' so beloved of some educators) also is seen as having a vital rôle. The social norms and frames of references within which learning tasks are set up, are recognized as more potent influences than, for example, serial position, in controlling the speed and direction of learning. Indeed from the standpoint of practical life the traditional learning experiment looks like a demonstration of the way 'trivial' influences can be made to affect improvement when all the more 'vital' influences have been eliminated from the experimental design.

It is true that for pure science, 'trivial' and 'vital' are irrelevant characterizations of the variables. An influence too small to be worth counting in practice may help to reveal what the basic psychological processes are. We do not, therefore, criticize the use of such variables as part of a theoretically-oriented experimental design. The error comes when we overestimate the value of such findings for the practical control of learning to the neglect of others more important. Specifically, in not one task in a thousand is such a typical laboratory variable as 'serial position' a material influence on the rate of learning. Even when the task is to learn unrelated items in a given serial order (auto licence

numbers, for example) our learning or failing to learn the item will far more often be determined by other factors than by where the item is placed in the series. For all that, the facts of serial position may be important for behavior theory and should be studied for that purpose. They are not much more than a footnote in applied psychology of learning.

For contrast, take the extended and careful studies of learning in arithmetic made by William Brownell. These studies are not suggested, so far as I can see, by any theory of learning nor can they be assimilated into any general theory of learning now current. Yet in my opinion they have revolutionary significance. Or consider the principles formulated by F. P. Robinson under the rubric of the higher study skills. They certainly do not derive from the traditional doctrines of learning and the more one examines the principles of higher study skills the more they appear as almost entirely independent of learning theory. Yet these empirically discovered principles make enormously more difference to the improvement of study performance than any application of what we know about serial position or forward association.

Or take the relationship between a child's learning and his liking for his teacher. We lack experimental evidence that liking for a person can directly influence the rate with which conditioned responses are formed, though some of Razran's evidence about the CR and eating together looks in that direction; and Mowrer's theory of two kinds of learning may yet throw light on this situation. The facts, however, are too complex to be put down as merely autonomic-based, emotional learning. Mowrer tells us how emotions get learned; yet to come is the theory and the evidence as to how emotion and the learning of geography interact. This, however, we know from direct observation; the child perceives, imagines, thinks and feels differently in respect to geography when he likes his teacher. And as a result, he learns differently, generally better.

The increased learning which comes from group membership must also be reckoned with. Which variable, do you think, makes more difference to the child's change in attitudes towards sex: the variable of the serial order in which the facts are presented; or the variable of teacher versus peer groups as the source of information? Or, if you don't like serial order, choose abso-

lutely any principle of learning based upon trying to make connections more efficient. Would you select frequency of repetition? But in the field of sex information we seldom employ repetition. I'll still stick at any rate to the social factors as making more difference. Or let me contrast Jost's law—one of the best established of connectionist principles—with a principle for which I can, more's the pity, as yet cite you very little experimental evidence.* I state it, however, in the form of a testable law. First, Jost's law, in case you have forgotten it: "When two associations or S-R bonds are of equal strength but of unequal age, repetition strengthens the older more than it does the younger." And now the proposed 'law of group identification': "A person will modify his attitudes, feelings and actions more quickly in the direction of conformity with those of a group with which he strongly identifies himself than in the direction of conformity with those of a group with which he weakly identifies himself."

Which, ladies and gentlemen, will you take as your guide when you next enter a grade-school classroom, the political arena or even your graduate seminar?

What I am trying to say is that our current learning theories, and probably any conceivable learning theories, are too restricted to cover the many facts which we must consider when we try to control or direct the improvement of behavior. The belief in the fundamental character of this or that learning construct puts blinders on us, keeps us from a flexible empirical approach to our technological problems, inhibits research because it is not 'fundamental,' and above all interferes with our finding the widest possible generalizations which will serve practical ends.

So I am urging first that pure theory will make more rapid advance if it abandons concern with the unfruitful construct of learning, and second that the term learning be returned to psychotechnology whence it came and where it is of use. Before psychology took it over, the verb 'to learn' was used in common speech to designate the whole complex process of reorganizing, improving, adjusting one's behavior. The term has its semantic rights and privileges to continue to be used in this inclusive

* T. M. Newcomb in *Social Psychology* (1950, p. 259) cites evidence from Lewin and Bales. Actually there is a great deal of factual evidence which suggests this theorem.

fashion. We must not try to confine learning within the straight-jacket of connectionism or of any other monolithic theory. To limit consideration of the factors which facilitate learning to those which directly facilitate or depress an alleged nuclear process of conditioning or connection-forming, is to sterilize them, to render them barren for the practical world from which the notion of learning originally came and to which it must ultimately return having issue as a result of its relations with psychology. A systematic technology of learning, however, freed from the necessity of trying to interpret miscellaneous facts in terms of a unitary scientific principle, can rescue the enormous range of facts which bear on practical problems, can put these facts in order for the use of teachers, foremen, and all others who in one way or another seek to help people improve their behavior.

If learning is thus a technological concept, does it belong to psychology? Not exclusively. There are considerations germane to learning which are non-psychological: philosophical, sociological, physiological or merely administrative. But the heart of learning as technology has always been and will continue to be psychological. To deal adequately with a practical problem of learning one may have to mobilize resources from the psychology of perception, or social norms, of motivation, of individual differences—almost anything, it may seem, but the 'psychology of learning' as usually conceived. To modify the attitudes and views of forewomen toward the employment of older women—that is, to get them to learn a new attitude—the technologist may need, as Marrow and French* have illustrated for us, the techniques of non-directive interviewing in order to discover the specific attitudinal barriers, may need the psychology of 'catharsis' in order to clear the emotional atmosphere, or the psychology of participation in order to insure a problem-solving approach. To insure retention of material learned in school, the technologist may need to find ways of providing that the material shall be ego-related for the particular learners.

These, clearly, are highly psychological principles; and the research which leads to their discovery and to the elucidation of their rôle in learning, the whole development of a systematic

---

* Alfred J. Marrow, and John R. P. French, Jr. "Changing a stereotype in industry." *J. of Social Issues*, 1945, 1, 33–37.

technology, is a task for the applied psychologist interested in actual learning activities for their practical value. That it will have fruitful relations with pure psychology goes without saying.

Such a technology of learning will make a direct attack upon significant practical problems. It will deal as closely as possible with the actual conditions under which the learner must work. Thus work with animals must be regarded as, at best, merely suggestive. A study of the effect of 'tuition' upon the way rats run the maze may or may not yield data for scientific theory; it yields practical conclusions only about how rats run mazes. Experiments with the pursuitmeter or with nonsense syllables have similar potential value for pure theory and, surely, similar limitations for application. If we wish to learn how children spell, as technologists we must study just that; we must conduct our experiments in schools, under the real conditions of school life, or as nearly so as possible.

I grant that such an approach has its limitations. The generalizations discovered lack 'spread' and, as indicated earlier, technological generalizations often overlap with each other in a way which for science would be exceedingly untidy and unparsimonious. When the organizing principle is that of use, there is apt to be much duplication and repetition.

But it is possible to bring order and system into the rules and principles of practical use in the field of learning just as it has been in engineering. The psychotechnology of learning will therefore have its theories. But they will be the theories which arise out of the problems of everyday living and they will be theories which lead us back into the field of application—i.e., where we help persons to reorganize their behavior for readier and more effective response to the conditions of life.

ACTION RESEARCH BY TEACHERS AND THE POPULATION SAMPLING PROBLEM[1]

STEPHEN M. COREY

Horace Mann-Lincoln Institute of School Experimentation
Teachers College
Columbia University

Action research has been variously defined,[2] but common to most definitions is the implication that this type of inquiry is undertaken by practitioners in order to guide and correct their present and future decisions and actions. The motivation for action research differs somewhat from the motivation for the traditional or fundamental educational research that we hear most about. Action research is undertaken on the hypothesis that the research approach to the solution of practical problems will result in better decisions and actions than will result if major dependence is placed upon intuition and subjective recall of evidence about consequences. The motivation for traditional or fundamental educational research, on the other hand, is to establish new generalizations and new truths of wide applicability which may guide the decisions and actions of other people as they try to cope with practical problems. Action research is undertaken by practitioners to improve their own decisions, and in the heat of combat, so to speak. Fundamental research is usually undertaken by students of education, in contrast to practitioners, and is usually characterized by considerable detachment from actual school operations.[3]

---

[1] The author is indebted to J. Wayne Wrightstone of the Bureau of Educational Research, Board of Education, of the City of New York, A. W. Foshay of the Horace Mann-Lincoln Institute of School Experimentation, and Helen Walker of Teachers College, for their critical reading of this manuscript. Each made a number of helpful suggestions. No one but the author, however, should be held responsible for the article as it appears here.

[2] See Isador Chein, Stuart W. Cook, and John Hardin, "The Field of Action Research." *American Psychologist*, Vol. 3, 1948, pp. 43–50.

[3] See Stephen M. Corey, "Action Research, Fundamental Research, and Educational Practices," *Teachers College Record*, Vol. 50, No. 8, May, 1949, pp. 509–514, and J. Wayne Wrightstone, "Research-Action Programs for Research Bureaus," *Journal of Educational Research*, Vol. 42, No. 8, April, 1949, pp. 623–629.

The most common analysis of the action research process as it applies to education suggests a pattern something like this:

1) In the course of their educational operations an individual, or a group, is troubled by a problem area in respect to which decisions and practices are judged to have been inadequate. Objective evidence descriptive of this problem area may or may not be procured.

2) In due course one or more rather specific and concrete problems are differentiated out of this large problem area because it is believed they need immediate attention.

3) On the basis of an examination of (a) their own past experience, and (b) the reported experience of others who have tried to cope with similar problems, the investigators hypothesize that if certain described and different practices are engaged in, the consequences will be more nearly in harmony with those that are desired.

4) These actions are put into effect, with careful attention given to a procedure for obtaining evidence that can be used to test their consequences.

5) This evidence is accumulated and examined carefully to see whether or not and in what degree it supports or rejects the action hypothesis.

6) Generalizations are inferred from this body of evidence and these generalizations are then used as a basis for future decisions and actions.

Because the practice of teaching or supervising or administering constantly involves problematic situations, this action research process is engaged in continuously as new problems arise. The quality of what the action researchers do in respect to each aspect of their inquiry may vary tremendously, not only from project to project, but within single projects.[4] This, however, is not unique to action research. The more traditional type of research investigations also vary greatly in quality. In either case, the confidence that can be placed in the results of inquiry is primarily a function of the quality of the steps that lead to these results.

One of the problems faced by action researchers is that their

[4] For a discussion of this quality problem see Stephen M. Corey, "Educational Research and the Solution of Practical Problems," *Educational Leadership*, Vol. IX, May, 1952, pp. 478–484.

findings are not necessarily applicable to other school systems or to school children in other school systems. The reason for this limitation is that the boys and girls or teachers who are the subjects of the action research do not represent a random sample of boys and girls or teachers in other school systems. This fact precludes extending more or less automatically the generalizations from action research studies to other populations which have not been randomly sampled in the studies. Teachers and supervisors and administrators in a given school system can with confidence put into practice the generalizations resulting from action research conducted in other school systems, only when the populations about which they are concerned are similar to those on which the action research was based.

An alternative and common sense justification for using what action research has revealed to be desirable in other situations is the subjective conviction, based upon the report of the action research itself, that one's own situation and that in which the inquiry was conducted are strikingly alike.

The population sampling problem that must be faced by people engaged in action research is implied by the nature of this type of inquiry. If a teacher in the third grade of a particular school building conducts action research, he does so for two reasons. First, as has been said, he wants to make better decisions and engage in better practices with the boys and girls he is teaching at the time. Secondly, he hopes that what he learns from his research to be a sound basis for action and decisions in connection with his present pupils will also be true for the boys and girls he will teach later on. The degree to which the generalizations from his action research studies in 1952 are applicable to future generations of his pupils in the same class and school is dependent, in a real sense, upon the extent to which his present pupils represent a random sampling from the larger cumulative population of children he will have in his classes next year, and the following year, and so on.

This implies that action researchers are bound by the same canons about generalizing which are emphasized, in theory at least, by those who engage in traditional educational research. The difference is that the action research generalizations are to be extended in a different direction. The fundamental researcher who conducts an inquiry using a random sampling of all of the

third grades in the State of Iowa or New York designs his study so that he will know the extent to which his generalizations are, in a sense, laterally applicable to all other third grades. The teacher engaged in action research, on the other hand, is concerned with the extent to which his generalizations can be extended vertically,—into the future,—and serve as guides for decisions and actions involving third-grade boys and girls he will teach but has not seen.

Whether or not generalizations from investigations undertaken by a teacher to provide a better basis for his decisions and actions in respect to his present pupils can be extended to his future pupils in the same school and at the same grade level depends upon the extent to which this hypothesis can be supported:

> Boys and girls in a teacher's particular class constitute a random sample from a larger cumulative population of his subsequent classes.

An analogous hypothesis of course could be formulated which referred to present and future populations of teachers or PTA's or any other groups in a specific locality.

This article reports several tests of the hypothesis that a present class can be thought of as a random sample from a total population of future classes taught by a particular teacher. A number of methods of testing were considered and two were selected. The major computations in each involved the following steps:

1) A critical variable was selected in respect to which test scores were available. The scores of pupils taught by the same teacher at the same grade level for a series of consecutive years were placed in a frequency table.

2) The frequencies in each class interval were expressed as percentages of the total cumulated population.

3) A frequency table, using the same class intervals as for the total population, was made of the scores for the critical variable for each of the separate classes.

4) Using these 'class' frequencies as observed data, their distribution was compared with a 'theoretical' or expected distribution resulting from multiplying the frequency in each interval for the single class by the interval percentage established for the total population (see 2 above).

5) The Chi Square test was applied to the null hypothesis that the distribution of scores in the individual classes and the distribution for the total population of N classes differed no more than might be expected as a consequence of chance or random factors.

These steps are illustrated below for the IQ's of three consecutive eighth-grade classes of pupils taught by a particular teacher in New York City.

STEPS I AND II

The total N for the three classes was ninety-six. Table I represents a frequency distribution of the ninety-six IQ's with the last column giving the per cent of the total number of cases falling in each class interval.

TABLE I.—FREQUENCY DISTRIBUTION OF IQ'S FOR TOTAL POPULATION OF THREE CONSECUTIVE CLASSES

IQ Interval	f	%
50– 69	1	1
70– 89	16	17.5
90–109	43	44.8
110–129	26	27.1
130–149	10	10.4
	N = 96	

STEPS III, IV, AND V

Using identical class intervals, Table II depicts in columns (1) and (2) a frequency table of IQ's for one of the classes. Column (3) is the 'expected' or theoretical frequencies derived from multiplying the N for this particular class (33) by the percentages in the last column of Table I. Columns (4), (5), and (6) merely represent calculations necessary for determining Chi Square, which in this case was 2.77.[5] With four degrees of freedom P

[5] The correction for continuity (see H. E. Garrett, *Statistics in Psychology and Education.* New York: Longmans, Green and Co., 1947. p. 246f) was not applied to these Chi Squares regardless of the small number of entries in some of the class intervals. The reason is that the uncorrected Chi Square is a more rigorous test of the null hypothesis and the latter was almost uniformly supported despite this added rigor.

TABLE II.—CHI SQUARE TEST APPLIED TO DIFFERENCE BETWEEN FREQUENCY DISTRIBUTION OF IQ'S FOR A SINGLE CLASS AND FOR TOTAL POPULATION

(1) IQ Interval	(2) Observed f	(3) Expected f	(4) (fo-fe)	(5) $(fo-fe)^2$	(6) $(fo-fe)^2$ fe
50– 69	1	.3	.7	.49	1.63
70– 89	4	5.8	1.8	3.24	.56
90–109	14	14.8	.8	.64	.04
110–129	11	8.9	2.1	4.41	.49
130–149	3	3.4	.4	.16	.05
	N = 33			$X^2 = 2.77$	
				df = 4	
				P = 60%	

becomes 60, which means that in sixty per cent of a large number of trials we would expect the distribution of IQ's for a single class to differ 'by chance' from the distribution of IQ's for the larger cumulative population by at least as much as the observed differences (Column 4).

The interpretation then is that the null hypothesis cannot be rejected. So far as intelligence test scores are concerned, this single class can be considered as a chance or random sample of the total three class population. This implies that the results of experiments conducted by this teacher, involving the pupils of this one class, and in which the IQ is a crucial variable, can be generalized to apply to his next two classes within the limits indicated by conventional formulas based upon sampling theory.

All in all thirty-eight similar Chi Squares were computed. Pupils in twelve different class sequences taught by twelve teachers in New York City and West Orange, New Jersey[6] were the subjects of the analysis, and the variables were IQ's and subject matter achievement and silent reading test scores.

The first analysis involved the IQ's of twenty-seven eighth-grade classes taught by nine New York City school teachers. Table III reports a frequency distribution of P's for the twenty-seven tests of the null hypothesis for nine sequences of three consecutive classes, each sequence taught by the same teacher.

[6] The author is indebted to Mr. Ronald Doll for making the West Orange data available.

TABLE III.—P's FOR TESTS OF NULL HYPOTHESIS APPLIED TO IQ's FOR EIGHTH-GRADE NEW YORK CITY PUPILS

Level	Frequency
80%–99%	13
60%–79%	4
40%–59%	5
20%–39%	4
10%–19%	1
	N = 27

In every case the null hypothesis could not be rejected. Any one of a series of three consecutive classes taught by the same teacher could be considered as a random sample of the total three class population so far as IQ's were concerned.

The second test of the hypothesis involved scores on the Iowa Silent Reading Test for three consecutive seventh-grade classes taught by the same teacher in West Orange, New Jersey, and four consecutive tenth-grade classes taught by the same teacher in that community. The number of Chi Squares possible to compute was seven and the P's were as follows: 95, 67, 62, 42, 42, 20 and 7. All of these P's are in excess of 5, which is the conventional magnitude for rejecting the null hypothesis.

The third test involved Standard test scores on the Language Arts and Arithmetic parts of the Stanford Achievement Test for four consecutive sixth-grade classes taught by the same West Orange, New Jersey, teacher. All of these Chi Squares were equal to or smaller than the number of degrees of freedom and the P's were 71, 53, 53, and 40.

Table IV summarizes the P's resulting from all thirty-eight

TABLE IV.—CHI SQUARE P's RESULTING FROM THIRTY-EIGHT TESTS OF NULL HYPOTHESIS

Level	Frequency
80–99	14
60–79	7
40–59	10
20–39	5
0–19	2

Chi Square tests of the null hypothesis. With the possible exception of 1, (% = 7), all imply that it would be hazardous to reject the hypothesis that the pupils in a teacher's class at any one time may, so far as the variables dealt with here are concerned, be considered a random sample from the larger population of several consecutive classes.

A second method of applying the Chi Square[7] test to these data involved no new computations but avoided the criticism that so large a fraction of the total population for any sequence of classes consisted of the population of the specific class which was being compared with this total population. Chi Squares were computed to test the null hypothesis that the distribution of frequencies of the classes in a particular sequence taken together do not differ more than might be expected by chance from the distribution in the cumulative population of these sequential classes (test of homogeneity). The P's resulting from these tests were 3, 15, 32, 37, 40, 47, 63, 61, 61, 68, 83, and 83. In only one case would it be statistically advisable to reject this null hypothesis.

CONCLUSION

The two analogous null hypotheses previously stated were supported substantially in the preponderant majority of instances when data were available for testing them. The only criterion used to get these test instances was the availability of scores over a period of years. It is conceivable that the pupil populations in a school might change so radically under some circumstances that no given population for one year could be considered as a random sample of a total population enrolled in the school over a period of years. It is also conceivable that small increments of change that are not appreciable for a three- or four-year period might become so over a longer period. Nevertheless, the analyses reported above provide considerable support for the practice of extending generalizations derived from action research studies of present populations to at least several future analogous generations within the limits suggested by conventional formulas based upon assumptions of random sampling.

---

[7] Suggested by Professor Helen Walker of Teachers College, Columbia University.

A VALIDATION OF DEVELOPMENT AND ADJUSTMENT HYPOTHESES OF ADOLESCENCE

AILEEN SCHOEPPE

New York University

and

R. J. HAVIGHURST

University of Chicago

THE PROBLEM

The concept of developmental tasks—those major common tasks which face all individuals in a given society or subgroup of society—is a useful tool for thinking about human development and about the education and guidance of children and youth. This concept seems to provide a framework within which knowledge about human behavior can be organized and this information can be used in learning optimum socialization processes. It has evolved from the recent efforts of students of the life sciences to understand social learning and the problems boys and girls face in becoming oriented to their cultural milieu. The developmental task concept originated in the Study of Adolescents of the Progressive Education Association, but was first elaborated in detail by Havighurst in 1948 in the monograph *Developmental Tasks and Education* (*3*). Since then it has been used rather widely. The most recent, thorough, and somewhat novel account is that by Tryon and Lilienthal in *Fostering Mental Health in Our Schools* (*7*). However, a complete survey of the literature shows it all to be theoretical.

The investigation herein reported was made in an attempt to learn more about the concept by studying intensively the achievement of thirty adolescents in Middle-western American society on five developmental tasks of adolescence; namely,

1) Learning an appropriate sex rôle.
2) Achieving emotional independence of parents and other adults.
3) Developing conscience, morality, and a set of values.
4) Getting along with age-mates.
5) Developing intellectual skills.

THE HYPOTHESES

The study tested several hypotheses about the relationships of levels of achievement on developmental tasks at various ages, and among several tasks at the same age. The aim was to answer the following specific questions: Does a tendency to be high in achievement in one task relate systematically to the level of achievement in other tasks at the same age? What are the patterns of achievement? Is there a systematic relation among the levels of achievement on the tasks at various periods of adolescence? If so, what is the nature of this relationship?

A concern for finding answers to the above stated questions led to the formulation of definite hypotheses which, when tested, would yield the desired information:

1) Good achievement on a developmental task at one age is followed by good achievement on similar tasks at subsequent ages.

2) Good achievement on a developmental task tends to be associated with good achievement on other tasks at the same age.

3) In a minority of cases good achievement on one developmental task may be used to compensate for poor achievement on other tasks.

As the study progressed, additional questions were raised. Chief among these were questions relative to the overlapping or discreteness of tasks. Specifically, are there basic underlying factors which may account for achievement in certain of the tasks and other basic factors which may account for achievement in other tasks? Does the 'clustering' of tasks influence achievement on them?

THE DATA

As a part of the Midwest Community Study Research Project, a major interdisciplinary research carried on since 1942 by the staff of the Committee on Human Development of the University of Chicago for the purpose of investigating character and personality development during childhood and adolescence in a typical Midwestern community, rich and extensive data have

been collected on a group of adolescents. These lent themselves to the purposes of the study and form the basic data for it.*

The subjects are fifteen boys and fifteen girls studied intensively in the Project. They are a selected sample, originally chosen on the basis of being 'adjusted' or 'unadjusted' as defined by arbitrarily determined composite criteria, from the original sample, which consisted of the one hundred fifteen children who had been born in 1932 and who in 1942 resided in the small city of 'Midwest' or its surrounding rural territory. The data used in the present study were compiled at ages ten, thirteen, and sixteen.

The completed folder of case-study materials on each subject, representing the results of thirty-two instruments, included information in the following broad areas: interviews; psychometric data; subject's reports via check lists; subject's reports via free response; various ratings by acquaintances, teachers, and others; projective techniques; sociometric data; physical data; Clinical Conference and Moral Character Conference† summaries and reports; ratings by staff members on a 'Trait Rating List' of forty-seven personality and social rôle items at age sixteen.

On the basis of these data eight or more Research staff members rated each subject on a ten-point scale on four developmental tasks; this scale represented a comparison by the rater of the subject against the Conference estimate of the norm for the American adolescent population with definitions set to place this norm at 5.5. At age sixteen separate ratings were made on the inner and outer aspects (i.e., covert and overt behavioral manifestations) of learning an appropriate sex rôle and assuming emotional independence of parents and other adults on the premise that these aspects were somewhat independent and might be discrepant, and that the data were sufficient to permit separate

---

* Without this previous work the present study could not have been possible; appreciation is expressed to all who worked on gathering and interpreting the primary data for the Midwest Community Study.

† The Clinical Case Conference studied the group from 1945 to 1947; it was followed by the Moral Character Conference. Both are part of the large Project and the latter integrated all the earlier data and their interpretation by the earlier Conference into its research.

ratings of them at this age. Ratings were also estimated from the accumulated data on these recurrent tasks at the earlier ages of ten and thirteen.* All of his ratings on a task were averaged for each subject.

In addition, it was thought other available data on the subjects permitted rating on a fifth task, 'Achieving intellectual skills.' In actuality, in the light of the data, this task would probably be more correctly termed 'Achieving academic skills,' for the two measures determining the ratings are measures of academic achievement only and, at least in our present society, 'intellectual skills' is a much more inclusive term and may have a connotation quite different from a narrow academic one. These ratings were not made by the Research staff, but were determined by decile rankings, made in terms of the entire original sample, on academic measures. For the school years ending in 1943 and 1946, this rating was made as follows: a numerical value was given to the average letter grade in each subject for the year, these numerical values were averaged, and the total scores of the entire sample ranked; this rank was then added to the individual's rank for the Metropolitan Achievement Test to give the total rank of the individual for the year; the total group for each year was then divided into deciles. For 1949 only school marks were available and these were in percentage form, rather than the letter grades of previous years; intellectual achievement for this year was rated by ranking these scores, then dividing the total group into deciles.† It is assumed these decile ratings thus determined are equivalent, for purposes of the present study, to the ratings on the other tasks on the ten-point rating scale.

---

* It is recognized that ratings for earlier ages probably would have been somewhat more refined had they been made at those ages rather than attempting to arrive at ratings for those ages simultaneously with ratings for age sixteen, but practical considerations in the Project made such great precision impossible. It is also recognized that it would have been highly desirable to have had such ratings of inner and outer aspects for all ages as were made for age sixteen, but the Research staff thought the data from which to rate adequately on inner feelings were insufficient at earlier ages to warrant a separate rating.

† The correlations between ranks for the total group were made by Spearman's rank-difference correlation formula and the critical ratio for significance determined. For 1943 data C.R. = 6.8; for 1946 data C.R. = 9.1. Thus there is a significant relationship between rank in the achievement test scores and rank in the school grades.

The ratings on the five developmental tasks, thus obtained, furnish the sole data for the correlational analyses to test the adjustment and development hypotheses.

METHODS OF ANALYZING DATA

To test the hypothesis of consistency of development, intercorrelations of ratings for all subjects on a task at each of the three age levels with each other age level were made to determine similarity of level of achievement of the task; this was repeated for all tasks.

To test the hypothesis of adjustment by considering simultaneous achievement on the several tasks, the ratings of subjects at each age level on all the tasks were intercorrelated to determine the degree of relationship between levels of achievement on the various tasks at given ages.

To check further on an adjustment hypothesis and the possibility of basic factors underlying certain tasks, intercorrelations were made of each task at all age levels with every other task at all age levels and these were organized by correlation profile analysis. In the same figure the profiles of congruent tasks at the same age were plotted. An attempt was then made by inference to determine an underlying factor which might be the causality for these task groupings.

To check on the hypothesis that good achievement on one task may be used to compensate for poor achievement on another or other tasks, rankings of levels of achievement were made for all individuals on all tasks at all ages. Those subjects who appeared simultaneously in the top quarter on one or more tasks and in the lowest quarter on one or more tasks were studied for a comparison of ratings, and an attempt was made to determine reasons for these discrepancies.

SUMMARY OF THE FINDINGS

Before considering relationships, it seems requisite to describe the central tendencies and dispersions of the group on the tasks at the various ages. Table I contains these data.

From this table it can be observed that the group consistently averaged near the theoretical mean (5.5) for the American adolescent population, as designated by the Conference definitions.

Thus, the group as a whole can be said to be, within the limits of the rationale of the study, quite average in their achievement on the developmental tasks rated, and the ranges indicate that the study includes cases from one extreme on the rating scale to the other extreme. Because of these wide ranges, opportunity is presented for studying all levels of achievement and for pointing up sharp contrasts.

TABLE I.—MEANS, RANGES, AND STANDARD DEVIATIONS OF RATINGS OF THIRTY SUBJECTS ON THE DEVELOPMENTAL TASKS AT THE THREE AGE LEVELS*

Task	Age	Mean	SE_{Mean}	Range	SD
SRO	10	5.73	.46	1.5– 9.4	2.54
	13	5.64	.41	1.5– 9.4	2.26
	16	5.83	.42	1.6– 9.9	2.32
SRI	16	4.47	.39	1.2– 9.6	2.15
EIO	10	5.03	.44	1.0– 9.8	2.42
	13	5.34	.42	1.0– 9.6	2.30
	16	5.70	.45	1.3– 9.8	2.46
EII	16	4.63	.44	1.3– 9.9	2.42
CMV	10	4.80	.45	1.6– 9.2	2.46
	13	5.04	.42	1.5– 9.4	2.32
	16	4.77	.42	1.8– 9.4	2.31
A-M	10	5.43	.56	1.2– 9.8	3.08
	13	5.27	.56	1.1–10.0	3.05
	16	5.53	.47	1.9– 9.7	2.58
IS	10	6.43	.57	1–10 decile	3.12
	13	6.47	.56	1–10 decile	3.09
	16	5.63	.54	1–10 decile	2.98

* The task abbreviations in the tables are: SRO—Learning an appropriate outer sex rôle. SRI—Learning an appropriate inner sex rôle. EIO—Achieving outer emotional independence of parents and other adults. EII—Achieving inner emotional independence of parents and other adults. CMV—Developing conscience, morals, and a set of values. A-M—Getting along with age-mates. IS—Developing intellectual skills.

Table II presents the basic correlational data which permit testing of the first two hypotheses and of the possibility of basic factors underlying certain tasks.

TABLE II.—INTERCORRELATIONS OF RATINGS ON EACH TASK AT EACH AGE LEVEL WITH EACH OTHER TASK AT ALL AGE LEVELS*

	SRO 10	*SRO* 13	*SRO* 16	*SRI* 16	*EIO* 10	*EIO* 13	*EIO* 16	*EII* 16	*CMV* 10
SRO 10	...	.61	.42*	(.34)	.68	.48	.50	.43*	.70
SRO 13	...	...	.84	.59	.62	.83	.84	.86	.38*
SRO 16	...	...	...	.70	.58	.71	.79	.75	.39*
SRI 16	...	...	...	...	.47	.76	.58	.68	.40*
EIO 10	...	...	...	...	...	.83	.79	.69	.43*
EIO 13	...	...	...	...	...	...	.95	.84	(.26)
EIO 16	...	...	...	...	...	...	...	.93	(.35)
EII 16	...	...	...	...	...	...	...	...	.47
CMV 10	...	...	...	...	...	...	...	...	...
CMV 13	...	...	...	...	...	...	...	...	...
CMV 16	...	...	...	...	...	...	...	...	...
A-M 10	...	...	...	...	...	...	...	...	...
A-M 13	...	...	...	...	...	...	...	...	...
A-M 16	...	...	...	...	...	...	...	...	...
IS 10	...	...	...	...	...	...	...	...	...
IS 13	...	...	...	...	...	...	...	...	...
IS 16	...	...	...	...	...	...	...	...	...

TABLE II.—*Continued*

	CMV 13	*CMV* 16	*A-M* 10	*A-M* 13	*A-M* 16	*IS* 10	*IS* 13	*IS* 16
SRO 10	.44*	.42*	.83	.57	.49	.45*	.42*	.37*
SRO 13	.53	.52	.53	.82	.77	.54	.49	.46*
SRO 16	.55	.53	.38*	.66	.73	.45*	.45*	.42*
SRI 16	.51	.49	(.28)	.44*	.48	(.34)	(.26)	(.18)
EIO 10	(.23)	(.22)	.57	.47	.39*	.37*	(.29)	(.23)
EIO 13	.38*	.37*	.45*	.64	.57	.51	.45*	.48
EIO 16	.50	.51	.51	.65	.63	(.30)	.55	.49
EII 16	.63	.65	.43*	.63	.64	.66	.62	.55
CMV 10	.80	.78	.80	.55	.45*	.57	.64	.58
CMV 13	...	.98	.60	.68	.78	.71	.78	.76
CMV 16	...	...	.58	.70	.78	.75	.79	.78
A-M 10	...	...	...	.75	.71	.62	.58	.52
A-M 13	...	...	...	...	.94	.69	.62	.58
A-M 16	...	...	...	...	...	.74	.72	.69
IS 10	...	...	...	...	...	...	.92	.78
IS 13	...	...	...	...	...	...	...	.91
IS 16	...	...	...	...	...	...	...	...

* An asterisk following a coefficient of correlation denotes it is significant only at .05 level; coefficients not so marked are significant at .01 level. A coefficient enclosed in parentheses indicates it is not significant at either .01 or .05 levels. With 28 degrees of freedon an r to be significant at .01 level must be .463 or greater; at .05 level .361 or greater.

DEVELOPMENTAL RELATIONSHIPS

The longitudinal study of relatedness of each task through the three age levels shows the hypothesis that good achievement on a developmental task at one age is followed by good achievement on similar tasks at later ages to be correct. In fact, all correlations except one are significant at the .01 level and that one—Sex Rôle (Outer) Task at ages ten, sixteen—is significant at the .05 level, that level of significance assumed as satisfactory for the study.

The one very clear-cut finding from the longitudinal study is the much greater variability among the tasks and the lower correlations for ages ten, thirteen than for ages thirteen, sixteen. The very high correlations at ages thirteen, sixteen show quite conclusively that the level of achievement on these particular tasks is practically fixed by age thirteen. Therefore, the period from ten to thirteen years seems the crucial period for adolescent changes and development in personality and socialization patterns in these subjects. A rather commonly accepted generality is that with puberty and physiological changes comes other adolescent development, but, if these correlations are valid, it appears that other adolescent changes may well forerun the physiological changes. There is need to extend the study to other earlier and also more advanced age ranges. But the findings suggest rather forcefully that the so-called 'latency' period may be a latency period only in physiological development and that it is a critical, extremely important period in social and personality development. Recurrently throughout the literature on personality development are suggestions that the latency period needs to be studied much more intensively than it has been; certainly these correlations hint that it and the early adolescent years are the time when permanent patterns are being formed and socializing influences are most effective. Therefore it behooves those guiding children during this formative period to have as much information as is possible on how the child may be aided in the accomplishment of the developmental tasks at this age period. An encouraging step in this direction is the very recent appearance of a new summary and critical analysis of available literature bearing upon the development and psychology of preadolescents, drawn from the various disciplines, with a formulation of guiding principles (*2*).

Some other conclusions regarding development may also be made. The correlations between age levels for Sex Rôle (Outer) Task are lowest; plausibly, physiological factors over which the individual exercises little control enter importantly into this task.

The correlations between achievement of intellectual skills at the various ages are highest; the factor of basic intellective capacity, over which the individual also exercises little control, enters prominently, but makes for stability instead of fluctuation on the task.

An extremely high correlation (.98) shows great consistency in the Conscience, Morals, Values Task between ages thirteen and sixteen; the conscience and morality pattern one has established by age thirteen evidently remains virtually stable unless there are traumatic or other situational experiences.

ADJUSTMENT RELATIONSHIPS

Overall Adjustment.—The hypothesis that good achievement on one task tends to be associated with good achievement on other tasks at the same age appears to be correct. Table III presents a summary tabulation of the correlations between achievement on the tasks at each of the respective age levels.

The adjustment, and also the development, hypothesis is further substantiated by the fact that seventy per cent of the correlations at age ten, eighty per cent at age thirteen, and ninety per cent at age sixteen are significant at the .01 level. Allport has succinctly described this consistency of personality: "As if to offset the disunity that comes with differentiation in early childhood, there is a compensatory process of integration. . . . . By virtue of the functional joining of psychical systems (through conditioning, generalization of habit, and all associational processes) integral units come into existence. For the most part these units represent coherent foci of development, found serviceable to adjustment and to mastery. . . . Functional units though to some extent independent tend, nevertheless, normally to converge into more embracing systems. Though perfect unity is never achieved, there may be said to be a constant progression in that direction." (*1*, pp. 344 f.).

A clear-cut conclusion is that satisfactory relations with peers is bound to accomplishment of the other tasks.

TABLE III.—SUMMARY TABLE OF CORRELATIONS OF TASKS AT VARIOUS AGES

	Age		
	10	13	16
Correlations Consistently High from Age to Age			
Sex Rôle (Outer) Task with Emotional Independence (Outer) Task	.68	.83	.79
Sex Rôle (Outer) Task with Age-Mates Task	.83	.82	.73
Emotional Independence (Outer) Task with Age-Mates Task	.57	.64	.63
Conscience, Morals, Values Task with Age-Mates Task	.80	.68	.78
Intellectual Skills Task with Age-Mates Task	.62	.63	.69
Correlations Consistently Lower from Age to Age			
Sex Rôle (Outer) Task with Intellectual Skills Task	.45	.49	.42
Emotional Independence (Outer) Task with Conscience, Morals, Values Task	.43	.38	.51
Emotional Independence (Outer) Task with Intellectual Skills Task	.37	.45	.49
Correlations Inconsistent from Age to Age			
Sex Rôle (Outer) Task with Conscience, Morals, Values Task	.70	.53	.53
Conscience, Morals, Values Task with Intellectual Skills Task	.57	.78	.78

Basic Factors.—From the correlation profile analyses, based on intercorrelations of each task at all age levels with each other task at all age levels,* the tentative conclusion is drawn that there may be three separate basic factors underlying the tasks studied, although much more definite scientific investigation

* The method duplicated that described by Tryon (*6*).

must be carried on before this can be stated with certainty. These findings are merely suggestive of the possibility of merit in a factor analysis of the tasks; such an analysis might do much to define, delineate, and delimit the specific tasks.

A factor which may be basic to the Sex Rôle (Outer) Task, Achieving Emotional Independence (Inner and Outer) Tasks, and Getting Along with Age-Mates Task has an affective base. These seem to center around the 'feeling' aspects of the individual—that part of his personality that, tragically, up to the present the schools have not generally deemed it their concern to develop. Physiological, emotional, and social maturity of the individual are all a consideration in this clustering, but the effect of each goes back to an emotional base and the individual's acceptance of his impulses.

A second factor which may underlie the Conscience, Morals, Values Task and the Intellectual Skills Task plausibly may be intellectual in origin. It suggests rational acceptance of discipline, or possibly a degree of conformity. This embodies intellectualizing by the individual and his acceptance of the necessity to formulate a pattern, a system of discipline, for his living both in informal interpersonal relations and in formal societal institutions.

A third factor seems basic only to the Sex Rôle (Inner) Task. About all one can definitely say from the data and the measures considered in the study is that this task does not fit with the others and therefore one deduces that it must have a different underlying factor. It is very tentatively suggested that this may have largely a biologic origin in hormonal balance. This hypothesis comes rather from using other studies to explain this discrepancy than from any data in the study. The isolation of this task corresponds with what Kinsey* seems to be concluding, and it is suggestive of Murphy's approach to personality study and his specific suggestion, with David Levy's study as evidence, that sexual and maternal drives suggest a constitutional factor (*5*, p. 106).

These are interesting data and may be valuable in suggesting approaches for further study to delineate in a more refined manner developmental tasks and to define them more precisely.

* Both in *Sexual Behavior in the Human Male* and from ongoing research on females.

But it must be reëmphasized that this is only an exploratory study suggested by the correlations resulting from the study of longitudinal development and cross-sectional adjustment on the tasks empirically defined by the Moral Character Conference and that more careful and extensive analyses must be made before accepting these as basic factors. However, this exploration does suggest that such research may prove fruitful.

Negative Relationships.—Tentative but positive conclusions were reached substantiating the third hypothesis—that some individuals use good achievement on one task to compensate for poor achievement on another or others. The small number of such cases (according to criteria, described previously, that were used for selection) makes any truly definite conclusions impossible but the trend suggests even a greater homogeneity in adjustment than anticipated. Only five of the thirty cases might be considered compensators, and only one subject appeared in both top and bottom quarters of achievement on tasks at all three age levels; thus compensation would seem to be for the most part and for most adolescents a temporary mechanism of adjustment to alleviate differences in physical maturation. It may be that compensation is used with expediency and forethought more often than hypothesized and is helpful in making a subsequent adjustment. Perhaps by the compensatory behavior the individual makes himself acceptable to himself during this particular period of development and, when the crisis has passed, tension is reduced and there is no further need to compensate in the same manner for the same reasons. However, this prognostication is wholly theoretical, may be invalid, or may be relevant only to periods of major physiological change such as adolescence. It is based primarily on empirical observation that rapid or retarded maturers physiologically often experience grave temporary maladjustments which time generally mollifies. The amount of compensation at stages of more stable bodily functioning should be studied for purposes of comparison and validation. However, the results do tend to indicate that achievement on one task at one particular age may be being used by an individual as a compensation for lack of achievement on another in order to make himself an acceptable self.

Finally, a study of the patterns of the five compensating cases shows that the patterns of negative relationships may be

similar, but for the same or totally different reasons. Two girls, while varying in actual ratings, have precisely the same patterns and the causes for their patterns seem to be similar; they rate high on Sex Rôle (Inner) Task but low on Relations with Age-Mates and Emotional Independence (Inner). Do they do this because they know that sex is an area in which they can achieve and they perceive that this will be socially acceptable behavior for females in our culture? Conversely, the achievements of a boy and a girl compensator form identical patterns but stem from very different reasons. The girl compensates neurotically, needing intellectual achievement to justify her proud isolation from good peer relations and acceptance of her sex rôle, while the boy seems to compensate because of immature development. These examples only serve to pose the question: what are patterns of compensation in our culture? They also serve to reiterate the need for studying individual cases and the dynamics of a particular situation.

SUMMARY

1) The results of this empirical study tend to show the hypotheses about adjustment and achievement implicit in the concept of developmental tasks to be correct. Some are so decisive as to leave little question; others suggest the need for further research to permit more definite conclusions.

2) The evidence seems clear-cut that the early period of adolescence is the crucial one in which changes in levels of accomplishment of these tasks are taking place, that levels of achievement are largely determined by age thirteen on these specific tasks.

3) Satisfactory relations with peers appear to be very closely linked to accomplishment of other tasks. Next in importance seems to be achievement of an appropriate sex rôle, and the findings indicate greater variability in achievement of this task at various adolescent age levels than variability in other tasks.

4) The analyses suggest the possibility of basic factors underlying the achievement of certain 'clusters' of tasks which may group or belong together because of these basic factors. Three such factors are very tentatively isolated.

5) Finally, the paucity of negative relationships serves to substantiate further the hypothesis of over-all adjustment and

to suggest that compensation may be to a greater extent than predicted a studied, oft times commendable means of temporary adjustment.

The study has raised questions and suggested further investigations. Similar comparable verification studies to test hypotheses for achievement and adjustment at other ages would be desirable. The possibility that there may be basic factors underlying the accomplishment of certain tasks which seem to cluster together needs further study, probably by factorial analysis. Finally, a much larger group of compensators must be studied through time before any very valid conclusions regarding negative relationships may be reached. Such studies should be fruitful in further defining the concept and giving it more specific meaning and perhaps wider application.

BIBLIOGRAPHY

1) G. Allport. *Personality: A psychological interpretation.* New York: Henry Holt & Co., 1937.

2) A. W. Blair and W. H. Burton. *Growth and development of the preadolescent.* New York: Appleton-Century-Crofts, Inc., 1951.

3) R. J. Havighurst. *Developmental tasks and education.* New York: Longmans, Green & Co. Rev. Ed. 1952.

4) A. D. Kinsey, W. B. Pomeroy and C. E. Martin. *Sexual behavior in the human male.* Philadelphia: W. B. Saunders & Co., 1948.

5) G. Murphy. *Personality: A biosocial approach to origins and structures.* New York: Harper and Bros., 1947.

6) Caroline M. Tryon. *Evaluation of adolescent personality by adolescents.* Monogr. Soc. Res. Child Develpm. Washington: National Research Council, 1939.

7) Caroline Tryon and Jesse W. Lilienthal, III. "Developmental tasks: 1. The concept and its importance. 2. Discussion of specific tasks and implications." In Caroline Tryon (Ed.), *Fostering mental health in our schools.* 1950 Yearbook of Association of Supervision and Curriculum Development, Washington D. C.: National Education Association, 1950.

THE FACTOR ANALYSIS OF SUBJECT-MATTER EXPERTS

WARREN S. TORGERSON AND BERT F. GREEN, JR.

Educational Testing Service and Princeton University

The application of factor analysis techniques to correlations between persons has often been suggested as a promising field for investigation. Stephenson[7] has been an optimistic proponent of this inverted factor technique and has developed the special case of Q-technique as a new approach to typology. Others, such as Thomson,[9] have expressed some doubts concerning the meaningfulness of correlations between persons, the criticism being that the scores added in computing such correlations may not all refer to a common dimension.

One application of inverted factor analysis which seems relatively free from this criticism is in the study of examiners' grades. If grades have been given to a number of individuals by a number of examiners, where all the examiners appraise the same product of an individual, we may consider the examiners as tests of the individuals or the individuals as tests of the examiners. Thomson[9] has discussed this problem, citing studies by Thomson and Bailes,[8] and by Burt[2] in which Spearman's factor technique was used in analyzing the correlations between examiners. Burt compared the Spearman and Hotelling techniques, calculating only one general factor. He discussed the possibility of group or overlapping specific factors, but ascribed little importance to them. It is, however, possible for group factors to be present, indicating that one group of examiners gave importance to elements not highly regarded by other examiners. It would be quite reasonable to expect such group factors to exist in certain areas of examination.

Closely related to the study of the nature of examiners' marks is the study of the validity of achievement tests. In considering this problem, it is necessary that a clear distinction be made between an achievement test and an aptitude test. An achievement test will be defined as a test which measures achievement in a specified area, and an aptitude test as a test which predicts future achievement. Since these definitions are in terms of the use and function of a test, it follows that any one test may be

both an achievement test and an aptitude test—a good measure of achievement in one area is often a good predictor of future achievement in the same or in a related area. The validity of the test as a measure of achievement, however, is not necessarily related to its validity as a predictor of future achievement.

The usual operational definition of validity as "the correlation of a test with a criterion," while reasonably adequate for determining the validity of an aptitude test, is inapplicable when considering measures of achievement. The achievement test is actually a measure of the criterion itself. This implies that the statistical techniques which have proved to be valuable in increasing the correlational validity of aptitude tests should not be applied uncritically to the achievement test field. The selection or rejection of items on the basis of the usual item analysis data is a case in point. While this method will very likely increase the correlational validity of a test, the effect on the validity of the test as a measure of achievement is doubtful.

Achievement test validity is perhaps best discussed in terms of the subject-matter area which a test is designed to measure. Useful in such a discussion is a concept of intrinsic validity, similar to that developed by Gulliksen.[4] If an achievement test is regarded as a sample of the universe of information, attitudes, ideas or skills that constitutes the subject-matter area, the test may be considered to possess intrinsic validity to the degree that it is a representative sample of that universe. We must rely on the judgments of experts in the subject-matter area to determine the extent to which a test is a representative sample.

If correlations between the judgments of a number of subject-matter experts can be obtained, a factor analysis of these correlations would furnish a representation of the structure of the experts' opinions. The use of factor analysis in this case is to study the dimensionality and complexity exhibited by the correlation matrix, rather than to discover any fundamental traits. If the experts were in agreement on the important variables in the area, then a common factor would be obtained. In general, the strength of the common factor would vary with the degree to which the same opinions were held by all of the experts. If differing schools of thought exist among the experts, then group factors would be found. The relative influence of each of these schools of thought would be shown by the size

and number of loadings on its group factor. In addition, the degree to which each expert stands alone in his ideas would be shown by the size of his specific variance. Thus, if one considers a single expert, his loading on a common factor would show the extent to which he agreed with everyone else; his loadings on the group factors, the extent to which he agreed with the various schools of thought; and his specific variance, the extent to which he agreed with no one but himself.

Such information on the structure of the opinions of the experts would be of great value in the construction of tests designed to measure achievement. To our knowledge, no such studies have been made.

Correlations between experts may be obtained for such an analysis in a number of ways. For an essay-type examination, such as an English composition, one could have the experts all grade the same set of essays. Correlations of these grades could be computed, considering the experts as variables, and the essays as population.

For an objective test one could have each expert construct a test according to his ideas of the important areas to be considered. These tests could then be administered to the same population, and a straight-forward factor analysis of the tests carried out. Alternately, one could use a Q-technique approach, wherein each expert would rate each of a large number of achievement test items according to significance for the variable being measured. Correlations of the experts could then be obtained by means of these ratings.

The present investigation is reported as an example of the use of inverse factor analysis in analyzing the judgment of experts. Although the study is limited by the nature and extent of the data, it is reported to illustrate the mode of attack proposed for such problems. The specific purpose of the study was to investigate the structure of opinion among readers concerning an English composition test. The presence and relative strength of a common factor was of primary concern, as well as the possible presence of group factors.

The data used in this investigation had been collected by the Statistical Analysis Department of the College Entrance Examination Board in an investigation of the reliability of scores on an English essay examination.[5] A sample of thirty-eight test papers

had been selected from responses to the essay question of the April, 1947, English Composition Test. The sample had been drawn so as to secure essays that varied widely in quality. Sixty English readers had independently read and graded these papers, reading a few each day over the entire period of examination scoring.

The total score on the essay was a composite of two part scores, a style score, and a material-and-organization score. Moreover, the former part score was composed of five subscores, and the latter, of three subscores. Scoring standards for these subscores had been used, based on the consensus of the entire group of readers. From these data an index of the 'reliability' of each reader had been computed. This index was a measure of the agreement of the reader with the other readers, taking into account differences in means, variances, and rank order of the assigned scores. The circumstances of the present study made it advisable to use a sample of about twenty readers rather than all sixty. Accordingly, a systematic selection of twenty readers was made such that the entire range of reader 'reliabilities' was represented. This method of selection was employed to insure as far as possible that the important dimensions existing in the entire group of readers would all be present in the sample. Since the reader 'reliabilities' are roughly comparable to the correlations of individual readers with the average of all the readers; i.e., the first centroid—this is equivalent to selecting a sample of tests from a larger battery so as to insure that the entire range of first centroid loadings would be represented in the sample. It was felt that this method would be likely to result in a matrix with the same common factor space as the entire group of readers.* The total score on the essay was taken as the unit to be examined.

The product-moment correlations between these readers were computed, resulting in a 20×20 matrix of intercorrelations based on thirty-eight cases. Four factors were extracted by means of Thurstone's centroid method of multiple factor analysis, after which it was judged that only random variation remained. Since one of the purposes in considering this example was to

* It could not, of course, result in adding any new dimensions. No method of selecting a sub-set of variables can result in a matrix of greater rank than that of the complete set.

obtain evidence concerning the existence of a group factor space, we must consider the hypothesis that one factor is sufficient to account for the correlations. With a small sample, there is no adequate statistical test of this hypothesis. In the present study, Tucker's,[1] (p. 125), Coomb's,[3] and McNemar's[6] criteria were applied; all indicated that one factor was not quite sufficient. Inspection of the first residual table also led to the subjective judgment that non-random variation remained, so enough factors were extracted to reasonably cover the possible subspace. The matrix of centroid factor loadings (F) is presented in Table 1.

TABLE 1.—CENTROID FACTOR MATRIX F

	I	II	III	IV	h^2
1	.89	−.21	.10	−.22	.89
2	.62	−.38	.20	−.20	.60
3	.76	−.30	−.13	.08	.69
4	.90	−.21	.23	−.05	.91
5	.76	.26	.06	.07	.65
6	.91	−.13	.11	.05	.86
7	.83	.07	.19	−.11	.74
8	.84	.16	−.22	−.18	.81
9	.88	.21	−.04	.11	.83
10	.92	.04	−.12	−.03	.87
11	.71	.07	.20	.18	.58
12	.89	.17	.06	−.12	.84
13	.88	−.08	−.22	−.08	.83
14	.66	.10	−.41	−.26	.68
15	.82	.29	−.07	.11	.77
16	.70	−.17	−.21	.30	.65
17	.88	.08	.18	.16	.83
18	.75	.26	.14	.07	.65
19	.89	−.22	.15	.14	.88
20	.83	−.10	−.23	−.08	.75

The reference axes were rotated by graphical procedures. It was possible to obtain positive manifold, but not a strong simple structure. Since it was discovered that one factor had high loadings for all people, this factor was arbitrarily set orthogonal to the other three, after the final rotations had been made. The rotated factor matrix (V) of projections on oblique axes is pre-

sented in Table 2, while Table 3 shows the transformation matrix Λ where $F\Lambda = V$. In Table 4 are the intercorrelations of the four factors.

TABLE 2.—OBLIQUE FACTOR MATRIX V

	A	B	C	D
1	.70	.48	.16	.21
2	.48	.58	.02	.00
3	.79	.25	−.01	.05
4	.73	.44	.28	.04
5	.58	−.04	.49	.24
6	.78	.29	.29	.08
7	.59	.25	.42	.22
8	.68	.05	.23	.51
9	.74	−.04	.44	.26
10	.80	.11	.26	.32
11	.59	.08	.44	−.01
12	.65	.15	.42	.36
13	.80	.17	.09	.34
14	.57	.01	.00	.59
15	.68	−.13	.46	.30
16	.80	−.01	.06	−.01
17	.73	.11	.48	.06
18	.55	.00	.53	.19
19	.81	.31	.26	−.06
20	.77	.17	.05	.32

The use of the rotational criteria of positive manifold and simple structure in an inverse factor analysis of this type requires further comment. The positive manifold restriction assumes, essentially, that the common factor space can best be described by achievement characteristics upon which the differences of opinion are differences in degree, but not in sign; that is, while any primary characteristic considered highly desirable by one expert may be considered of little or no importance by another, it is not likely to be considered by another as a negative characteristic which should be penalized. The opinions of experts in most achievement areas may be expected to satisfy the positive manifold restriction.

However, after accepting this restriction, there will remain,

TABLE 3.—TRANSFORMATION MATRIX Λ

	A	B	C	D
I	.85	.19	.33	.25
II	−.25	−.68	.70	.48
III	−.31	.46	.59	−.50
IV	.36	−.54	.21	−.68

TABLE 4.—INTERCORRELATIONS OF OBLIQUE FACTORS R_p

	A	B	C	D
A	1.00	.00	.00	.00
B	.00	1.00	.26	.16
C	.00	.26	1.00	.06
D	.00	.16	.06	1.00

ordinarily, a large number of possible orientations of the reference vectors. The primary interest is in obtaining reference vectors which will adequately represent the common factor space and which are interpretable in terms of achievement variables. If rotation is to the best simple structure, interpretation of the factors will be facilitated. It is important to note, however, that a strong, overdetermined simple structure is not necessary in inverse factor analysis—and, very likely, cannot even be obtained if the experts are selected so as to secure a representative sample of expert opinion in the achievement area. Whether or not the simplest structure obtained from another sample of experts from the same population is identical with the first is not crucial. It is only necessary that the common factor space be the same in each case, and that it be possible to describe this space in terms of the reference vectors chosen. Rotation to the simplest structure, then, is useful because it tends to place the vectors in those positions where they are most easily analyzed.

In the present study, the rotated factors have an interesting pattern. The interpretation of Factor A, the general factor, is relatively clear. Since the loadings on this factor are all high, the readers were in general agreement concerning the quality of the essays. This, of course, is of primary importance if the scores on the test are to be consistent. Furthermore, the general factor loadings were for the most part higher than the group factor loadings, indicating that the consistent differences of opinion represented by the group factor space were of secondary

importance in determining the essay scores. It should be remarked that, if the restriction of a positive manifold is granted, the first centroid cannot be directly interpreted as an indication of the degree of general agreement. It is theoretically possible that in rotating to an orthogonal positive manifold, the general factor will disappear. In such a case, one would have to conclude that no general agreement existed. It is true that with loadings on the first centroid as high as those obtained here it is impossible to rotate to an orthogonal group factor structure without a general factor. Nevertheless, the size of the loadings on the general factor after rotation is a better indication of the degree of general agreement than is the size of the first centroid loadings. In the present situation, the first centroid accounts for eighty-eight per cent of the common variance or sixty-seven per cent of the total variance, while Factor A accounts for sixty-four per cent of the common variance or forty-eight per cent of the total variance.

The small group factors indicate that to some extent there were consistent differences of opinion among the readers. Because of their low magnitude, relative to the general factor, their interpretation is difficult.

For each group factor, the three or four readers with high loadings on that factor and similar low factor loadings on the other group factors were selected. For each of these readers, the variances and intercorrelations of the part scores given to the essays were computed. The correlations and variances of readers high on one group factor were compared with those of readers high on the other group factors. It was found that for each reader high on Factor B, the intercorrelation of the major part scores—style, and material-and-organization—was higher than that of any of the C or D factor readers. The average intercorrelation for the three B factor readers was .68 as compared with an average value of .46 for the four C factor readers, and of .37 for the three D factor readers. Thus Factor B readers tended more to mark the essays on the basis of one impression, while the others tended more to differentiate between material-and-organization and style. Factor B, then, is at least partly the result of a halo effect.

Further study of these selected readers met with no success. Their subscore and part-score variances showed no consistent

group patterns. A subjective examination of those essays graded either higher or lower by one group of readers than by each other group did not reveal any differentiating characteristics.

There may, however, be a sex differential on two factors. It was noticed that the six readers with the highest loadings on Factor C were men, while six of the seven readers high on Factor B were women. (Readers 1, 3, 4, 6, 7, 12, 13, 14, and 19 were women.) Several hypotheses might be advanced concerning the reasons for this apparent sex difference. Preferences for certain types of essay content or certain stylistic preferences might be expected. However, no aspect of the data was found which helped to explain this sex difference. It remains a possibility for further studies to explore.

No further interpretation of the group factors was possible. Since the analysis was based on only thirty-eight cases, it is quite possible that Factor D represents largely error variance, and that error variance and true variance are confounded in the other group factors.

The main conclusion which may be drawn from this pilot study is that in the situation under investigation, there was considerable over-all agreement among the readers or experts. In addition, the small group factor space indicates that to some extent readers tended to separate into groups such that within group agreement was greater than that between groups. Since the effect of the prepared scoring standards was undoubtedly to increase the general factor at the expense of the group and specific factors, it is probable that different schools of thought do exist in this field, although their effect has been minimized by the scoring procedure.

This example of the use of inverse multiple factor analysis is of course limited by the small sample and by the grading restrictions necessarily placed on the readers. The method, however, when applied on a larger scale to the judgments of experts in subject-matter fields should be of great value in the construction and evaluation of standardized measures of achievement.

REFERENCES

1) Robert Blakey. "A re-analysis of a test of the theory of two factors." *Psychometrika*, 1940, 5, 121–136.

2) Cyril Burt. "The analysis of examination marks." In P. Hartog, and E. C. Rhodes. *The Marks of Examiners*. London: MacMillan and Co., 1936.

3) Clyde H. Coombs. "A criterion for significant common factor variance." *Psychometrika*, 1941, 6, 267–272.

4) Harold Gulliksen. "Intrinsic validity." *Amer. Psychologist*, 1950, 5, 511–517.

5) Edith M. Huddleston. "Recent studies of the English composition test." *College Board Review*, 1948, 1, 45–54.

6) Quinn McNemar. "On the number of factors." *Psychometrika*, 1942, 7, 9–18.

7) William S. Stephenson. "The inverted factor technique." *Brit. J. Psychol.*, 1936, 26, 344–361.

8) Godfrey H. Thomson and S. Bailes. "The reliability of essay marks." *The Forum of Education*, 1926, 4, 85–91.

9) Godfrey H. Thomson. *The Factorial Analysis of Human Ability*. Boston: Houghton Mifflin Co., 1948.

SEQUENCE IN MULTIPLE CHOICE ITEM OPTIONS*

SCARVIA B. ANDERSON

George Peabody College For Teachers**

One of the basic aims of test construction is, or should be, that the subject's test score be determined by his knowledge of the subject matter of the test rather than by any shrewdness in finding mechanical aids in the test. In multiple choice items even the order of the options may furnish subjects with systematic mechanical devices which will enable them to 'beat' the test. For example, many authors, often unconsciously, have preferences for placing correct responses in certain positions; and on some tests the lengths of options may be related to their correctness, when the correlation between length of option and correctness should ideally be zero. On the other side of the problem, certain arrangements of options may make an item more difficult than is desirable with a given situation or subject matter.

Mosier and Price (*2*) in 1945 presented a method of assisting the test constructer in overcoming his biases in arranging five option multiple choice items. In preparing a table of random order for such items, they wrote the permutations of the numbers 1 through 5 in cyclic order and assigned each permutation a sequence number from 1 through 120. These sequences were then assigned to the table according to the order in which each sequence number occurred among the last three digits of a nine-place logarithm table.

Table I presents a similar approach to randomizing five-option multiple choice item options when randomization is desirable. After writing the items and editing them, the test constructor may start at any place in the table and proceed in either a vertical or a horizontal direction without skipping any permutations. For example, if he were to begin with the sequence in the first row and first column, he would label his first option of Item I 'C,' the second option of this item 'B,' the third 'A,' and so on. Then if he went on vertically, the first option for Item 2 would be 'E.'

* The writer is indebted to Dr. Julian C. Stanley for helpful suggestions during the preparation of this article.

** The author is now associated with Tufts College, Systems Coördination Project, Naval Research Laboratory.

TABLE I.—120 RANDOM PERMUTATIONS TO BE USED IN ARRANGING THE OPTIONS OF FIVE-OPTION MULTIPLE CHOICE ITEMS

CBAED	BDECA	ADECB	ADCBE	DCABE	EBACD	ACBDE	EBADC
EADBC	ACDBE	BDEAC	CEABD	DCAEB	DACBE	CDAEB	DCEBA
EDBAC	CDBAE	DEBCA	DCEAB	DBCAE	DCBAE	BEACD	DEBAC
CEBDA	CAEBD	ECDAB	BADEC	DEACB	DBACE	ACDEB	ECADB
AECDB	ABEDC	ABDEC	DCBEA	BCEAD	EADCB	CBEDA	CABDE
EDCAB	EBDCA	ECBDA	ADBCE	BCDAE	EDBCA	CEADB	CBDEA
ECABD	EBDAC	CBDAE	DABEC	CADBE	ABDCE	BDAEC	EACDB
BEDCA	DEABC	EDABC	ACEDB	CADEB	BAEDC	BCADE	DBAEC
ECBAD	BDCAE	BACED	BAECD	EABDC	BCDEA	EBCDA	DECAB
DBECA	ABCED	ADEBC	DECBA	CDABE	DAEBC	BEADC	BEDAC
DABCE	CEDAB	CBEAD	EBCAD	BCAED	DACEB	BCEDA	CDEAB
CDEBA	AEDBC	CAEDB	CEBAD	EDCBA	AEDCB	BECAD	ADCEB
ABCDE	EABCD	ECDBA	EDACB	BDACE	EACBD	AECBD	BDCEA
DAECB	CBADE	AEBDC	CEDBA	AEBCD	BECDA	ABECD	DBEAC
BACDE	ACEBD	ACBED	BADCE	CABED	ADBEC	CDBEA	DBCEA

If the test constructor is interested merely in randomizing the correct option for each item without regard to the order of the distracters, he should list this option first for each item and then use the first letter of each permutation to determine proper positions (or the second letter of each, or the third, or the fourth, or the fifth). However, he should not use the first letter of a permutation for Item 1, the second letter of the same permutation for Item 2, and so on, since this procedure would destroy randomness.

This table, in the writer's opinion, has two decided advantages over the Mosier-Price table. First, the answers are presented as letters instead of as numbers. Many test writers prefer letters for the options, in order that the complete items may be numbered without confusion. In addition, standard machine-scored answer sheets usually have letters instead of numbers for the individual choices. Second, the sequences in this table are presented in random order. The permutations in cyclic order were assigned randomly to the table by the use of Tippett's *Random Sampling Numbers* table (*3*), while Mosier's and Price's use of logarithmic tables as a method of securing true randomization may be questioned (see Kendall, *1*, p. 193).

In the five option multiple choice item it is obvious that there are only five possible places for the correct answer, so that the 120 permutations would have 'A' in the first place twenty-four times, in the second place twenty-four, and so on. In studying

the sequence of the correct option from item to item, it can be seen too that one can occasionally get a correct response sequence such as AAA or BBBB, the greatest possible run on a letter being twenty-four in one hundred twenty questions.

Apparently no really crucial experiment has yet been done concerning the influence of the placement of distracters upon the item itself. Such a study seems to be in order, for if, contrary to popular opinion, the arrangement of the decoys makes no difference, all the test writer needs to do is randomize the correct options according to the table of Mosier and Price or the table here.

As a natural corollary of 'How to randomize' arises the question 'When not to randomize.'

Mosier and Price (*2*) have brought up the situation in which there is a 'best' answer and a 'nearly best' answer to an item. One of their studies indicates that the better students do not read all options of an item but mark the 'nearly best' choice if it comes in a position before the 'best' choice. They suggest that such an item discriminates against the better students and that in such an instance the 'best' option should precede the 'nearly best' by one or more positions.

The same authors indicate that date responses should be presented in chronological order or, alternatively, by a method of modified randomization. The table could be consulted for the random position of the correct response, and then the incorrect responses could be devised and grouped around it in chronological order by the test constructer.

The following items represent various situations in which the writer feels that randomization should not be used:

1) *Item having two or more similarly worded options.*

Example: Underlying the calculation of a product moment coefficient of correlation is the assumption of

A independence of measurement of the two variables.
B normality of the distribution of both variables.
C rectilinearity of regression of each variable on the other.
*D equality of the means of both distributions.
*E equality of the variances of both distributions.

2) *Item having a 'none of these' or 'all of these' response* (see Wesman and Bennett, *4*).

Example: In which of the following respects is a long test necessarily superior to a short one?

A Reliability
B Validity
C Objectivity
D Representativeness of sampling
*E None of these

3) *Item in which certain options are 'A,' 'B,' 'C,' 'D,' or 'E.'* (Such items may be found in the word "fluency" section of the *American Council on Education Psychological Examination,* when subjects are asked to mark the first letter of the word which answers the item.)

Example: The father of individual mental testing was

A G
*B B
*C C
D P
E S

4) *Item in which certain options are numbers corresponding to possible positions among the options.*

Example: The arithmetic mean of the X distribution, M_x, is 4. The arithmetic mean of the Y distribution, M_y, is 8. $n_x = 8$; $n_y = 4$. Approximately what will be the grand mean when the two distributions are combined with $N = 12$?

A 6
B 8
C 12
*D 4
*E 5

5) *Item having numerical responses, where certain incorrect choices are not arbitrary but clearly indicated by the data.* (When the choices, other than the correct one, are more or less up to the test constructer's imagination, the Mosier-Price, *2,* method of modified randomization suggested for dates, might be utilized.)

Example: On a test with a standard deviation of 10 and a mean of 50, an individual with a raw score of 75 will have a z score of

A	−2.5	or	A	25.0
B	0.4		B	2.5
C	1.5		C	1.5
D	2.5		D	0.4
E	25.0		E	−2.5

Two other possible plans of arrangement of multiple choice item options are alphabetical arrangement and arrangement according to the length of the option. Alphabetical arrangement by the first letter in the first word of the option is particularly applicable when the responses are proper names. Arrangement according to the printed length of the option, of course, may involve using the shortest option first and the longest option last or vice *versa*. It is felt that both of these plans should be experimentally compared with that of random arrangement.

SUMMARY

The elimination of test-constructer biases in the order of multiple choice item options is viewed as an important problem in test construction. To assist the test writer in randomizing the order of five-option multiple choice item responses a new table, which the author feels has definite advantages over an earlier table by Mosier and Price (*2*) is presented. In addition, several suggestions are made concerning situations in which random arrangement of options is probably unsuitable.

BIBLIOGRAPHY

1) Maurice G. Kendall. *The advanced theory of statistics,* Vol. 1. London: Chas. Griffin & Co. Ltd., 1948.

2) Charles I. Mosier and Helen G. Price. "The arrangement of choices in multiple choice questions and a scheme for randomizing choices." *Educ. psychol. Measmt.*, 1945, 5, 379–382.

3) L. H. C. Tippett. "Random sampling numbers." *Tracts for computers,* No. xv. London: Cambridge Univ. Press, 1950.

4) Alexander G. Wesman and George K. Bennett. "The use of 'none of these' as an option in test construction." *J. educ. Psychol.*, 1946, 37, 541–549.

BOOK REVIEWS

EDWIN S. SHNEIDMAN AND OTHERS. *Thematic Test Analysis.* New York: Grune and Stratton, 1951.

The book which Shneidman and his collaborators offer is a highly specialized illustration of the use of projective material. One schizophrenic patient was tested with TAT, Shneidman's MAPS, and other instruments. The protocols for the thematic tests were given to fifteen experts for blind analysis. The intent was to show various systems for tabulating, inspecting, or extrapolating from thematic records. The participants who exposed their blind analyses to comparison with the life history were brave, and fortunately are also highly competent. The thematic tests were treated by Arnold, B. Aron, Bellak, L. Eron, and others even better known. In addition to the analyses and interpretations of TAT, the book contains a wise introduction by Murray, a description of both tests, analyses of Rorschach, Wechsler, MMPI, Draw-a-Person, and Bender-Gestalt protocols, and full case history plus therapy notes. Finally, we have a synthesis.

The book is best regarded as a file of raw material of great interest to the investigator who wishes suggestions for dealing with protocols, and to the advanced student of projective methods and abnormal psychology. The many interpreters demonstrate the wide range of viewpoints now existing in psychodiagnosis. The range from objectivity to impressionism is fairly represented, and the writers of the separate chapters make clearer than does Shneidman that the different systems are intended to answer rather different questions. There is an inevitable tendency to compare the writers in terms of the adequacy of their guesses vis-à-vis the case history. To conclude that X's method did not lead to comprehensive insights into the patient is manifestly unfair, if X's system of analysis is intended for a specific research use. The interpreters who tried to describe the full dynamics of the case seem to have been rather successful. The record selected includes a great deal of frankly paranoid material, and generous verbalization—but is not an unreasonably easy one to interpret.

The negative criticisms that can be made, granting the intent of the book, are few. The inclusion of the MAPS test is understandable, but of seemingly minor value. Most of the interpreters made some gestures toward MAPS interpretation, but in most cases the treatment is perfunctory and adds little to the TAT. Whether MAPS could add more if scored by those accustomed to it is an open question; but this book would be tighter knit and equally useful without MAPS.

The interpretations of supplementary tests are of varying value. Rorschach is thoroughly and coherently treated by Klopfer. The Wechsler and MMPI materials are routine. The analyses of Bender-Gestalt by Hutt and of Draw-a-Person by Machover must be characterized as astonishing. The Bender calls for drawing simple geometric figures from memory. Hutt tiptoes step by step over some very tenuous analysis to the judgment that the root difficulty is a dominant mother figure toward whom the patient is very attracted sexually. That this fact is verified by the case history makes the reasoning process no less remarkable. Machover is equally free to elaborate on the smallest of indicators, and again her interpretations would be dismissed as pure fancy were it not for the correspondence with the case history. It would surely be dangerous to suggest to students that they can interpret projective tests as freely as these experts do; but it is well to have this cleanly gathered material at hand as an example of the ideal of the diagnostician. Under negative remarks, we note also the bad printing of some pages.

The final value of this book will depend on the ingenuity with which it is used. Despite Shneidman to the contrary, it is not 'a research.' He promises some later statistical analyses, but it seems unlikely that formal analysis will elicit any worth-while generalizations. (To ask which method works best, for example, is not feasible in a study where $N = 1$.) The material is first and last food for thought. One can envision some fine seminars where this book will be used with students who can be both imaginative and critical. The book lends solid support to present enthusiasm for thematic techniques. But what ideas one takes away from the book will depend on how one reads it, for the book has no central message.

LEE J. CRONBACH
University of Illinois

WILMA DONAHUE AND DONALD DABELSTEIN, Editors. *Psychological Diagnosis and Counseling of the Adult Blind.* Selected Papers from the Proceedings of the University of Michigan Conference for the Blind 1947. New York. The William Byrd Press, 1950, pp. 173.

Scant and lacking in diversity is the professional literature about the blind. This is the assumption expressed by the representative for the American Foundation for the Blind which joined with three agency participants in publishing the proceedings of the Conference on the Psychological Diagnosis and Counseling of the Adult Blind held at the University of Michigan in 1947. The twelve papers presented in this volume are, with one single exception, taken from contributions made at this conference. The selection was made with the purpose in mind of widening the horizons of the workers in the field for work purposes as well as to delineate research problems and methods fundamental to the improvement of psychological and rehabilitation services for the adult blind. Topics covered in these papers include: Somatopsychological aspects of blindness; mental hygiene; community and family problems; employer attitudes; measuring intelligence, mechanical and manual ability of the blind; the use of projective techniques and the Minnesota Multiphasic Personality Inventory for the blind; and vocational counseling.

Many facts based on observations and experimental findings as well as attitudes are expressed in the volume by the twelve contributors. Not all of them are equally valid. Some of the contributors neglected facts of personality dynamics as well as facts of individual differences and the blind were characterized as if all were alike. Most of the contributions, however, are more critical as well as more clinically valid. Some facts considered are certainly deserving of consideration and have significance for occupational and social adjustments of the blind as well as the society in which they live. Facts presented in the chapter on employer attitudes are disheartening and definitely indicate the need for education of employers. In this chapter we are informed that of the entire blind population comparatively few secured employment in wartime and approximately fifty per cent of those who became employed lost their jobs after the

surrender of the enemy. The fact that agencies receive testimonial letters from the employers is not changing the picture. Instead, it indicates the need for educating employers to practice what they preach.

Some of the more important results from psychological studies of the blind in the first half century that have been established are considered here. Progress has been slow. The hope is expressed that the developments in research techniques in recent years in things proving useful in other psychological domains if applied to problems of blindness might promise to be of equal value. The contributions in this volume can be definitely helpful to workers in the field who wish to be informed. For research orientation the supplementing of this volume with the larger and more ambitious volume on *Blindness—Modern Approaches to the Unseen Environment*, edited by Paul Zahl, is suggested by the reviewer.

H. MELTZER

Psychological Service Center
St. Louis, Mo.

NATHANIEL CANTOR. *Dynamics of Learning.* Second Edition. Buffalo: Foster and Stewart, 1950. $3.00.

Those who have been expecting major revisions in the second edition of Nathaniel Cantor's *Dynamics of Learning* will be disappointed. No re-working of the materials is evident, although a summary chapter has been added to Part II. With this exception the book remains what it was in 1946, a vigorous attack on traditional education, and a brilliant demonstration of the learning process, showing psychological principles in operation in the classroom. Not least is an informing sense of the crisis of our age and a faith in newer educational methods for developing more truly democratic living in our society.

With its aims the present reviewer is in complete agreement, and they are as fully pertinent today as when the book first appeared. The volume is in no sense out of date. The writing is as stimulating and alive as ever. Yet one might have expected somewhat sharper outlines to appear in the new edition, some more important additions, a tightening of the relations of the concepts to one another, and perhaps some refinements in the application of these concepts. Although normally out of place

in a review, it may be worth while on the present occasion to illustrate in some detail a number of such possibilities, leaving aside the difficult and highly controversial question of the scientific validity of the psychological principles which Cantor uses.

If one examines Chapter 4, dealing with the basic principles, it seems that Cantor's postulates are (1) that human behavior always involves conflict or ambivalence; (2) that the human organism resists changing itself; and (3) that paradoxically the human organism is adaptive, and seeks constantly for adjustment and growth. These postulates are similar to those of other dynamic theories of personality, but unlike the formulation by O. H. Mowrer and Clyde Kluckhohn, for example, in their *Dynamic Theory of Personality*, Cantor does not make explicit use of the situation or 'field.' It seems to be implied in the third postulate, but the emphasis is placed on 'will.' In contrast to Theodore M. Newcomb's study of *Personality and Social Change*, Cantor's treatment is principally intra-personal and tends to slight the influence of school or community, of group or clique, on rôles and attitudes of students. This omission, by a sociologist, is rather surprising. For completeness, it would seem desirable that this aspect of group behavior be included in Cantor's learning theory.

In dealing with ambivalences, Cantor regards two sets as basic. One of them is the conflict of dependence and independence, the other a conflict between positive and negative 'will.' Both express the dual nature of the 'self.' In respect to these concepts there are two principal difficulties. One is that the definitions are somewhat confusing. On the one hand they are described as patterns of behavior of individuals organized toward particular ends. The intra-personal character of conflicts is not kept clearly separate from inter-personal relationships. Without warning Cantor shifts from the intra-personal to the operational level, sometimes within the same sentence. While it may be true that these two aspects or processes may occur simultaneously and are certainly interrelated, it seems preferable that they be kept separate analytically.

The second and more important problem is the existence of two sets of ambivalences, both of them 'polar' conflicts. Why two sets? If they are identical, one set would certainly be adequate.

If they are not, then the differences should be clearly described and their theoretical reference explained. The distinctions, however, are nowhere to be found. Superficially, the sets seem identical; actually they seem to differ in meaning and to constitute distinct systems. One set seems to conform more to concepts in social psychology and may be derived from the writings of James Plant, the other resembles the theories of Otto Rank. Whatever their derivation, each set has a number of advantages which Cantor has been able to exploit for descriptive purposes, but inevitably the coherence of his system suffers. It may be said, though, that when one looks at the details, as worked out in the second part of the book, it seems as if the conflict of positive and negative will, with its accompanying concept of guilt, is predominant. It is this conflict that seems to be related most closely to other concepts, such as resistance, projection, and identification. The responsibility for pointing out such theoretical applications, however, clearly lies with the author, not the readers. The second edition might well have eliminated the use of the dual, or perhaps duplicate, systems.

Similar difficulties arise with the principles elaborated in Part II, especially resistance and projection. Resistance, illustrated in a separate chapter, is defined as the refusal of an organism to change. One assumes that the new term will refer to a distinct principle, since concepts used earlier are not mentioned, but actually it seems identical with negative will, the postulate that 'fundamentally, the organism resists changing itself.' The term 'resist' does occur in the latter definition, but hardly justifies its sudden use as a substitute for negative will, unless there is at least a word of explanation.

A further overlapping occurs with the term projection, together with other confusions. For one thing, projection is commonly used in psychology to designate the defense mechanism in which a person refers his own impulses, thoughts, feelings or wishes to another individual or sometimes to an object, thereby denying the reality of his own disagreeable thoughts and feelings. Cantor, on the other hand, defines it as a process by means of which an individual rejects any experience which threatens the self as previously organized. There is some similarity in meaning between these definitions, yet they describe processes operating in different theoretical systems. Apparently they resemble one

another just enough to cause misinterpretation and considerable misunderstanding.

As Cantor uses the term, projection is related to negative will. In situations calling for change or adjustment, a person will "refuse to attend to the challenge or avoid it or modify it so that it fits in with his established habits. He may also meet the challenge of change by reaffirming his positive. Whichever alternative is employed, what he is doing, essentially is asserting himself. He is projecting his negative will." Projection, in this sense, is the active principle of negative will. It is a construct referring to a variety of reactions, all of which are assumed to arise out of a refusal of an individual to alter his organized self. The 'refusal' to change, or the general tendency of every person to maintain his previously-established organization, is negative will.

Although this usage does some violence to the literal meaning of projection ("to throw forward or outward") when it is applied to several kinds of avoidance behavior and to self-depreciation, which becomes a 'projection inwards,' the fact is that the application of the word is of limited usefulness. It does refer to the operational aspects of negative will, but its focus remains primarily on the negative function—the preservation of the *status quo* of the organism. The result is practically a tautology. There is no real gain in precision or elaboration. It would seem more logical and more helpful to proceed from negative will to a classification of types of behavior to which it gives rise when a person feels threatened or is called upon to change.

If, for example, a number of students remain in silent and stubborn opposition to the behavior expected in a classroom situation, one might label their reaction 'rejection.' If the response takes the form of strong affirmation or insistence on a person's own point of view, it might be called 'assertion.' If certain students are self-depreciatory or wish to avoid a problem by seeking refuge in their past experiences, their behavior could be named 'self-criticism' or 'flight.' Persistent and unwarranted criticism or a continuous complaint response might be termed 'attack.' A mechanical response such as superficial restatement of text materials might be designated 'conformity.' When the materials are misinterpreted or modified so as to fit a person's established attitudes or habits, the process could be

called 'distortion.' All of these and other categories, then, would describe common types of behavior designed to preserve intact the person's 'self.' One or some combination of them would probably be adopted by students on different occasions, although some degree of consistency in selection might be predicted. All of them would very likely be supported by rationalizations if they were challenged. At any rate, they are all manifestations of negative will, and are the kind of responses which presumably prevent 'learning' from occurring.

If Cantor had developed similar types of classification, his book would have gained greatly in precision and in usefulness in the classroom, and it might lend itself more easily to scientific testing. There is some recognition of such possibilities in his chapter on projection, for he does develop considerably one specific kind of reaction—that which we have called 'assertion.' Unfortunately, he names it 'projection,' so that in his usage the term stands not only for a variety of responses which actively represent negative will but also for a particular one of this group. This practice causes great confusion, which might have been avoided by setting up the kinds of categories we have suggested.

It is worth noting that in respect to positive will, which presents the same kind of problem in theoretical development, Cantor does differentiate three 'levels of learning' in his chapter, "The Student's Will-To-Learn." The first is a superficial intellectual level represented by students who select and repeat with little change statements from the text or whose comments garble the text and show little or no comprehension. The second is the level of thinking through the implications of the data, and the third is the level achieved by those who not only see the conceptual implications of the material but also apply it in a way that involves more of their total self. Each of these levels is illustrated by a series of excerpts from student papers. These have the virtue of illustrating written reactions in their complexity, but this method also tends to prevent the more complete formulation of the different levels and to strain the interpretive powers of readers.

When one examines these 'levels,' moreover, one finds that the first is inappropriate. It represents a manifestation not of positive but of negative will. It seems to refer to the types of response which we have suggested be called 'conformity' and

'distortion.' They are exactly the kinds of reaction which Cantor repeatedly maintains cannot be regarded as genuine learning. If this is correct, then we are left with only two categories or 'levels' of learning, with some question as to whether the one is 'higher' or 'better' than the other, as seems to be implied.

Again it would seem desirable to differentiate more or less distinct types, which may or may not lie on a continuum. Let us suppose, for example, that one early expression of positive will in the classroom would be 'questioning,' which is essentially a device of orientation. It may reflect a defensive attitude but more probably a genuine curiosity, so that a skillful teacher could guide an interested student toward seeking answers to his questions largely by his own efforts, thereby encouraging in him self-learning. Another response might be called 'discovery,' referring to those students who find in the text statements which crystallize their own feelings or attitudes, or which suggest new insights or a sense of expanding horizons. Understanding is implied and some degree of new confidence. 'Exploration' might be used to characterize the reactions of those students who examine the reading material critically, who investigate related areas of knowledge, who seek further support for the ideas and concepts under consideration, and who uncover unsolved problems. Cantor's second level seems in part to apply to this kind of development.

Corresponding to Cantor's third level is the type of response one might label 'application.' It is found in those students who are able to perceive the relevance of the course materials either to experiences of their own or to the behavior of others. Their work often shows a generous use of personal data or observations which are simultaneously a demonstration (though often indirectly) of the accuracy of their understanding of the concepts and of their insight into situations or events that are not specifically dealt with in the test or in the discussion. Their reactions show a generalizing power that finds its expression in the concrete rather than the abstract, in life contexts rather than in the realm of knowledge.

Another category might be designated 'cognition.' Cantor seems to mix it with his other levels, but it might profitably be isolated. Although it may grow out of the processes we have

referred to as 'exploration' and involve some degree of 'application,' every instructor knows students who are fascinated by the nature of the concepts that are being studied. They are more aware of the theoretical aspects of the materials, the kind of framework or system that is being presented, and more likely to make comparisons of similar or different systems, or to analyze the 'logic' of the concepts. Even the best undergraduates probably cannot go far in these directions, but their growth can be observed in the degree of their penetration and in the range of their grasp.

Whether 'cognition' is better or is less desirable than 'application,' with its 'life' orientation, depends upon one's point of view. Perhaps there are basic differences in personality and interest which may make the use of a continuum of 'learning,' unless within specific types, a highly dubious procedure. Certain students may restrict themselves and confine their academic work largely to one type or to a particular combination of responses; others may show distinct differences of development at various points in their lives. In this respect the present reviewer again feels a strong inadequacy in the *Dynamics* for not including broader social, community and group contexts as important factors in the motivation, the focus and the type of learning responses of students.

In addition to developing classifications to indicate various types of behavior which characterize negative and positive will, Cantor might well do the same with another key concept, identification. No one seems to have formulated or analytically differentiated types of identification, or marked off degrees of intensity, if these are important to learning. What is the effect, for example, of the instructor's becoming a substitute parent for students? What if he exists as a generalized authority figure, is strict or indulgent? Does the 'permissive' teacher often fail because students have no effective models to emulate in terms of self-learning? What is the result of peer identification in the classroom as compared with focusing on the instructor? Tentative answers to these and related questions might well find their appropriate place in future editions of the *Dynamics*.

Finally, a few words may be said about the summary which Cantor provides in his second edition. Its absence from the first edition left readers without any clear-cut conception of the

essential 'dynamics' of learning. In a sense the whole book stood—and still stands—as a simultaneous description and illustration of the process, and readers often had the feeling that the process had been demonstrated. But usually they were unable to translate these feelings into a more exact statement of these dynamics. My own formulation of them prompted Cantor, as he acknowledges, to supply his own version for the present edition, and it is certainly a helpful addition. It is also influenced to some extent by the criticisms which are presented in this review. The will construct, for example, is emphasized as fundamental, and the materials on projection are omitted. Important as these modifications are, their full value cannot be realized by confining them to a summary chapter. At present they remain, in fact, somewhat inconsistent with the treatment in the earlier chapters, and they cannot become wholly effective until they are worked into the body of the book.

Readers of this review will no doubt recognize my essential approval and deep appreciation of the contribution which Cantor has made in his *Dynamics of Learning.* Out of this valuation has grown the hope that he will extend and develop his original ideas so as to build a more solid foundation for the understanding of inter-personal relationships within and outside the classroom that will further the self-development of students and foster more democratic ways of living and thinking in our society.

MILTON C. ALBRECHT

The University of Buffalo

HECTOR CHEVIGNY AND SYDELL BRAVERMAN. *The Adjustment of the Blind.* New Haven: Yale University Press, 1950, pp. 320.

The premise of this book is that the problems of the blind can be understood best through a psychoanalytic study of the attitudes toward blindness held by the blind and by the sighted. The authors develop this theme cogently and effectively, capitalizing on their different backgrounds: the senior author's experience as a blind person and the co-author's experience in the practice of clinical psychology.

The proposition is offered that the blind are treated by society as a minority group and encounter difficulties similar to those met

by other minority groups, e.g., Negroes and Jews. In many respects this book is a polemic against the misconceptions and prejudices about the blind held by the sighted and by many of the sightless, against the view that the blind should be handled as a group separate and apart from other handicapped groups, and especially against the misguidance and selfish interests that have entered into the efforts of many of those who have attempted professionally to assist the blind. The history of our own society and of other societies in developing conceptions of blindness and in caring for the blind is traced in an effort to substantiate the authors' contentions.

The usefulness of the psychoanalytic approach in obtaining an understanding of the problems of blindness is interestingly demonstrated. Man's overestimation of vision and consequent underestimation of the other senses are held responsible for much of the error in judging the potentialities of the blind for learning and adjustment. The overvaluation of sight in the sexual area, the sexual significance of blindness, and the mobilization of fears related to the castration complex by the thought of blindness are discussed in a clear and provocative exposition. The motivation behind types of adjustment popular among the blind and behind varieties of treatment given to the blind by their sighted acquaintances is examined from a psychoanalytic standpoint. Some grim conclusions are drawn from this analysis.

The authors make a strong plea for the 'integration' approach in dealing with the blind. This means that the blind will be encouraged and permitted to take their places in society as individuals and not as members of a group or as specimens of a type. This is considered the only approach consistent with a democratic philosophy. It is pointed out that the success and failure of the blind in attaining this goal has paralleled the rise and fall of the democratic spirit in societies.

There is much of interest and value in this book, not only for those concerned with the blind, but also for anyone interested in the psychology of vision, minority group problems, or the problems of handicapped persons. It should appeal to the general reader not only because of the issues involving everyone, which are discussed in an interesting and stimulating manner, but also because of the charges it makes against society, and the challenge it offers to reëxamine one's attitude toward the blind. To special-

ists in personality adjustment and social psychology, the book demonstrates the enriched understanding that can result from a skillful application of psychoanalytic principles to problems in their areas. This book deserves a wider reading public than its title will probably attract.

PHILIP M. KITAY
Adelphi College

PUBLICATIONS RECEIVED

Dorothy C. Adkins and Samuel B. Lyerly. *Factor Analysis of Reasoning Tests.* Chapel Hill, N. C.: The University of North Carolina Press, 1952, pp. 122. (paper) $2.00.

Charlotte Buhler, Faith Smitter, Sybil Richardson, and Franklin Bradshaw. *Childhood Problems and the Teacher.* New York: Henry Holt & Co., 1952, pp. 372. $3.75

Alvina Treut Burrows. *Teaching Children in the Middle Grades.* Boston: D. C. Heath & Co., 1952, pp. 280. $3.75.

Clyde M. Campbell, Editor. *Practical Applications of Democratic Administration.* New York: Harper & Brothers, 1952, pp. 325. $3.00.

Raymond B. Cattell. *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist.* New York: Harper & Brothers, 1952, pp. 462. $6.00.

George S. Counts. *Education and American Civilization.* New York: Bureau of Publications, Teachers College, Columbia University, 1952, pp. 491. $3.75.

James Drever. *A Dictionary of Psychology.* Baltimore (3300 Clipper Mill Road): Penquin Books Inc., 1952, pp. 316. (paper) $0.85.

Andre Fouche. *La Pedagogie des Mathematiques.* Paris, France: Presses Universitaires de France, 1952, pp. 134. (paper)

Norman Frederiksen and W. B. Schrader. *Adjustment to College: A Study of 10,000 Veteran and Nonveteran Students in Sixteen American Colleges.* Princeton, N. J.: Educational Testing Service, 1951, pp. 504.

John W. French. *The Description of Aptitude and Achievement Tests in Terms of Rotated Factors.* Chicago: The University of Chicago Press, 1951, pp. 278. (paper) $4.00.

Milton S. Girvitz. *The Dynamics of Psychological Testing.* New York: Grune and Stratton, Inc., 1951, pp. 396. $6.75.

Phyllis Greenacre. *Trauma, Growth and Personality.* New York: W. W. Norton & Co., Inc., 1952, pp. 328. $4.50.

Ruth E. Hartley, Lawrence K. Frank, and Robert M. Goldenson. *Understanding Children's Play.* New York: Columbia University Press, 1952, pp. 372. $3.50.

Ernest V. Hollis and Alice L. Taylor. *Social Work Education in the United States.* New York: Columbia University Press, 1951, pp. 422. $5.50.

George Humphrey. *Thinking: An Introduction to Its Experimental Psychology.* New York: John Wiley & Sons, Inc., 1951, pp. 331. $4.50.

Lloyd A. Jeffress, Editor. *Cerebral Mechanisms in Behavior: The Hixon Symposium.* New York: John Wiley & Sons, Inc., 1951, pp. 311. $6.50.

Henry Enoch Kagan. *Changing the Attitude of Christian toward Jew: A Psychological Approach through Religion.* New York: Columbia University Press, 1952, pp. 155. $2.75.

Louis Kaplan and Denis Baron. *Mental Hygiene and Life.* New York: Harper & Brothers, 1952, pp. 422. $3.50.

G. Marian Kinget. *The Drawing-completion Test: A Projective Technique for the Investigation of Personality.* New York: Grune and Stratton, Inc., 1952, pp. 238. $6.75.

Raymond G. Kuhlen. *The Psychology of Adolescent Development.* New York: Harper & Brothers, 1952, pp. 675. $5.00.

Hannah Logasa. *Historical Fiction.* Philadelphia: McKinley Pub. Co., 1951, pp. 280. $4.00.

Elly McKean. *It's Mine.* New York: The Vanguard Press, Inc., 1951, pp. 40. $2.00.

Margaret Mead and Frances Cooke MacGregor. *Growth and Culture: A Photographic Study of Balinese Childhood.* New York: G. P. Putnam's Sons, 1951, pp. 223. $7.50.

William C. Morse, Francis A. Ballantime and W. Robert Dixon. *Studies in the Psychology of Reading.* Ann Arbor, Mich.: University of Michigan Press, 1951, pp. 188. (paper) $2.50.

Morris Ploscowe. *Sex and the Law.* New York: Prentice-Hall, Inc., 1951, pp. 310. $3.95.

Otto Pollak and Collaborators. *Social Science and Psychotherapy for Children.* New York: Russell Sage Foundation, 1952, pp. 242. $4.00.

Arthur Robson. *The Eternal Truths of Life.* Adyar, Madras, India: The Theosophical Publishing House, 1952, pp. 296.

Klara G. Roman. *Handwriting: A Key to Personality.* New York (333 6th Avenue): Pantheon Books Inc., 1952, pp. 382.

Ephraim Vern Sayers. *A First Course in Philosophy of Education.* New York: Henry Holt & Co., 1952, pp. 399. $3.75.

Douglas E. Scates, Bernard C. Murdoch and Alice V. Yeomans. *The Production of Doctorates in the Sciences: 1936–1948.* Washington: American Council on Education, 1951, pp. 228. (paper)

Saul Scheidlinger. *Psychoanalysis and Group Behavior.* New York: W. W. Norton & Co., Inc., 1952, pp. 245. $3.75.

Alexander A. Schneiders. *The Psychology of Adolescence.* Milwaukee: The Bruce Pub. Co., 1951, pp. 550. $4.00.

Helen Shacter. *Understanding Ourselves.* Bloomington, Ill.: McKnight and McKnight Pub. Co., 1952, pp. 124. (paper) $0.70.

G. Milton Smith. *More Power to Your Mind.* New York: Harper & Brothers, 1952, pp. 180. $2.50.

Ross Stagner and T. F. Karwoski. *Psychology.* New York: McGraw-Hill Book Co., 1952, pp. 582. $5.00.

C. P. Stone, Editor. *Annual Review of Psychology: 1952.* Stanford, Calif.: Annual Reviews, Inc., 1952, pp. 462. $6.00.

Hilda Taba, Elizabeth Hall Brady, John T. Robinson, and William E. Vickery. *Diagnosing Human Relations Needs.* Washington: American Council on Education, 1951, pp. 155. (paper) $1.75.

Edwin Forbes Tait. *Textbook of Refraction.* Philadelphia: W. B. Saunders Co., 1951, pp. 418.

George C. Thompson. *Child Psychology: Growth Trends in Psychological Adjustment.* Boston: Houghton Mifflin Co., 1952, pp. 667. $5.50.

Willard F. Tidyman and Marguerite Butterfield. *Teaching the Language Arts.* New York: McGraw-Hill Book Co., 1951, pp. 433. $4.50.

Roy DeVerl Willey. *Guidance in Elementary Education.* New York: Harper & Brothers, 1952, pp. 825. $5.00.

Karl P. Zerfoss, Editor. *Readings in Counseling.* New York: Association Press, 1952, pp. 639. $6.00.